Optimal Grading

Zubrickas, Robertas

Abstract: Assuming that teachers are concerned with human capital formation and students - with ability signaling, in this paper we model a teacher-student relationship as an agency problem with conflicting interests. In our model, the teacher elicits effort from a student rewarding for it with a grade, the utility of which to the student is the ability signal inferred by the job market. In the event that the job market does not observe individual teachers' grading practices, teachers find grades as costless rewards and optimally choose to be lenient in grading. As a result, "the problem of the commons" of good grades emerges leading to the depreciation of grading standards and grade inflation. The prediction of the model that the lower the expectations the teacher holds about her students' abilities, the flatter the grading rules she sets up is empirically supported.

DOI: https://doi.org/10.1111/iere.12121

Originally published at:
Zubrickas, Robertas (2011). Optimal Grading. In: 26th Annual Congress of the European Economic Association, Oslo, 25 August 2011 - 29 August 2011.
DOI: https://doi.org/10.1111/iere.12121

Optimal Grading*

Robertas Zubrickas[†]

October 8, 2010

Abstract

Assuming that teachers are concerned with human capital formation and students — with ability signaling, in this paper we model a teacher-student relationship as an agency problem with conflicting interests. In our model, the teacher elicits effort from a student rewarding for it with a grade, the utility of which to the student is the ability signal inferred by the job market. In the event that the job market does not observe individual teachers' grading practices, teachers find grades as costless rewards and optimally choose to be lenient in grading. As a result, "the problem of the commons" of good grades emerges leading to the depreciation of grading standards and grade inflation. The prediction of the model that the lower the expectations the teacher holds about her students' abilities, the flatter the grading rules she sets up is empirically supported.

Keywords: Principal-agent model, teacher-student relationship, costless rewards, grading rules, mismatch of abilities and grades, grade inflation, teacher incentives.

JEL codes: C70, D82, D86, I20.

1 Introduction

Contrary to what is expected, grading criteria for assessing student performance are invariant neither over time nor across universities or separate study fields. There is a number of stylized facts about grading patterns documented in the literature on educational measurement, and in this paper we study two of them, arguably, most notable ones. First, there is a tendency for grading standards to depreciate over time—commonly known as the "grade inflation" phenomenon. Second, professors apply more stringent grading criteria in fields with more able students, and *vice versa*, leading to a mismatch or low correlation between students' grades and their abilities. (Empirical evidence is extensively

*I would like to thank Tore Ellingsen, Christian Ewerhart, Drew Fudenberg, Magnus Johannesson, Karl Wärneryd, and seminar participants at the University of California, Santa Barbara.

[†]University of Zurich, IEW, Winterthurerstrasse 30, CH–8006 Zurich, Switzerland. E-mail: robertas.zubrickas@uzh.ch.

discussed later in the text.) In addition, to take a more general stance, we discuss the compression of ratings and leniency bias phenomena from the literature on performance evaluation, which can be intricately linked to the two grading patterns above.

The main purpose of this paper is to rationalize the aforementioned grading patterns from the perspective of an individual teacher, who looks for a grading rule to achieve the benevolent objective of increasing student knowledge in her taught subject. The crux of the arguments presented here is centered on particular aspects of grades as teacher rewards for students: their costlessness and asymmetry in information about grading rules applied. In our model, the grading rule optimally designed by a teacher and its dynamics resultant from changes in distribution for student abilities and in perception of grade value offer a good match to the grading patterns empirically observed. Though this paper is mainly concerned with the positive analysis of optimal grading and its properties, here we also discuss some normative aspects of the questions raised.

The approach of this paper is to consider a teacher-student relationship as a principal-agent model with conflicting interests: a teacher's goal is to pass on knowledge to students, who only care for the teacher's assessment of their performance—grades. A justification of this conflict of interests would be the dichotomy of the role of education: human capital formation versus job market signaling.[1] An interpretation of the modeling framework to be presented is that a teacher is concerned with the human capital formation side, while a student—with the ability signal his or her accomplished education carries along, treating knowledge obtained as abstract and useless outside academia.

In the model, the teacher (principal) offers her students a "contract," which is a grading rule assigning grades to exam scores. It is assumed that the teacher has a technology—the exam test—that allows her to assess her students' knowledge levels attained, on which she conditions grades to be rewarded. With the help of grades, which cost nothing to reward, the teacher aims to elicit costly learning effort from—equivalently, to pass on knowledge to—her students, who come from a population of students with disparate abilities for the subject taught. We assume that a knowledge level attained by a student is in a direct and deterministic relationship with his learning effort elicited (conditional on his ability type). The main constraint that we impose on the teacher is that her grading rule needs to be incentive compatible, i.e., a grade cannot be conditioned on the student's ability level, which is his private information (but the distribution for abilities in population is known to the teacher).[2] The crucial element of the model is the value that students attach to the grades rewarded.

[1] See, e.g., Bedard (2001).

[2] If the hidden-information framework seems restrictive—arguably, teachers have access to students' previous records and can learn about their abilities—then, alternatively, we could require that the teacher cannot discriminate among her students by applying ability-specific grading rules (i.e., the same exam score has to result in the same grade irrespective of the student's ability level). With this alternative formulation, the optimization problem would remain intact as in the case with the hidden-information framework adopted, and, therefore, the latter is retained for its link with the existing literature.

We separately consider two sources of grade value for students: 1) nominal—a grade is of a value on its own—and 2) relative—the value of a grade comes in the form of an ability signal inferred by the job market. The latter interpretation of grade value is in the tradition of treating education as job market signaling (Spence (1973)). Under this interpretation, we also distinguish between two situations i) the job market can observe the grading rule applied by a teacher, and ii) the job market cannot observe the grading rule applied and makes its inference about a student's ability based on its own perception of grading standards applied. In situation i), an individual teacher has to take into account the effect the stringency of her grading rule has on the grade value perceived by students, whereas in situation ii) this effect is absent, essentially making grades valued by students nominally.

With the nominal interpretation of grade value, the key assumption embedded is that in a given class of students an individual student's utility of a grade is independent of the (expected) distribution of grades in the class. We can think of several justifications of this assumption. Besides the one already suggested—the ability-signaling value of a grade is independent of the class distribution of grades when the job market does not observe the grading rule applied—another justification would be that students can be ignorant about the ability distribution in the class or they are of bounded rationality to discern the resultant grade distribution from the grading rule set up by the teacher. Another rationale behind modeling grades nominally would be that a student may care for his grade at every class he takes not just because of its immediate job-market-signaling value but for its contribution toward his final grade average, which later on will serve as an ability signal. Finally, many external criteria of academic achievement, used for, e.g., scholarship application purposes, are usually assessed in nominal grades.

Therefore, in the case with the nominal value of grades—where, to put it differently, there is no interdependence of (*ex ante*) utilities between any two students in a class and, as a result of that, there is no scarcity of rewards, i.e., grades to give—the teacher-student model turns alike a *single-agent* (i.e., single-student) agency problem with hidden information or a monopolistic screening problem. In particular, the teacher designs an incentive-compatible grading rule to maximize the student's expected knowledge, where the student cares for his grade only. However, there is an important difference with the canonical monopolistic screening problem: in our model, the reward is *costless* to grant but of a value to the student. In other words, in our model the transfer (grade) function does not enter the teacher's objective function but only the student's, implying that there is no intercomparison of utilities. In the case with the relative value of grades and observable grading rules, where the teacher faces a scarcity of grades, the single-agent framework still applies after modifying the student's utility function of grades, as shown later in the text.

In the static setting with a continuum of types, we obtain the following predictions about optimal grading patterns. In the event when students value grades nominally, the teacher pools student ability types for the highest grade (of an institutionally preset grading scale). This result arises because it cannot be optimal for the teacher to reserve the highest grade to one ability type because of its small probability mass. By turning more lenient with high-ability students, the teacher can extract more effort from lower-ability types, and this gain in effort from lower types outweighs the corresponding loss of effort from the highest types (unlike in the standard model with costly transfers). In other words and more generally, the "no pooling at the top" property does not hold when the principal can costlessly reward the agent.

For the case with the relative value of grades, if the student's learning costs are not too high (in the model, the effort cost function is not too convex), then, unlike in the nominal-value case, the teacher's optimal grading rule perfectly screens student ability types.[3] However, if the teacher's grading rule is not observed by the job market, we again obtain pooling student types for the highest grade but even at a larger scale than in the nominal-value case. In addition to the reasons for pooling types discussed above, here we also observe the deterioration of the job-market perception of grades as ability signals, eventually leading to the depreciation of grade value for students and, as a result, to teachers' pooling types even further.

Next, our comparative statics analysis shows that in the model with nominal grades if the teacher holds low expectations about her students' abilities, then she should apply more lenient grading rules (in order to elicit on average higher learning effort), and *vice versa*. This can lead to heterogeneous distributions of grades among classes different in student abilities—in particular, to a mismatch and low correlation between students' grades and their abilities.

Significantly, the existing empirical evidence strongly supports the findings of the model, lending credibility to our chosen modeling strategy of a teacher-student relationship. With regard to the comparative statics result, Goldman & Widawski (1976) report a negative correlation between students' Scholastic Aptitude Test scores (which could be seen as a proxy measure of students' abilities) and the grading standards in the classes the students were majoring in. According to this study (conducted at University of California, Riverside), the negative correlation observed is due to the fact that professors in a field consisting of students with high abilities tend to grade more stringently than do professors in a field with lower-ability students—precisely as our model predicts. These empirical findings were confirmed by similar studies conducted at Dartmouth College (Strenta & Elliott (1987)) and at Duke University (Johnson (2003)). With regard to the pooling-at-the-top result, more generally, our model is consistent with the compression of ratings and leniency bias phenomena about raters' shallow differentiation of performance

[3] This result is in line with the results from related models modeling non-pecuniary rewards such as status incentives; see Moldovanu *et al.* (2007), Besley & Ghatak (2008).

and tendency to over-rely on the top end of the rating scale (see, Murphy & Cleveland (1995)). Later in the text, we discuss empirical evidence more thoroughly.

If taken in a dynamic perspective, our model also offers an insight into the grade inflation phenomenon—"an ongoing rise in grade point averages without an accompanying rise in student ability or effort.[4]" Within the model, we can identify two factors contributing to the grade inflation. First, it is the change in the distribution for student abilities toward the lower-ability end.[5] According to our model, with larger numbers of less able students in classes teachers set up less stringent grading rules, which translate into higher average grades (and less study effort[6]). As already mentioned before, the reason for it is to extract more effort from increasingly numerous lower-ability types (even at the expense of further distorting incentives for high-ability types). The second factor is about the opacity of grading practices or rules arising due to an increasing number of university openings available.[7] With the number of educational institutions increasing, an individual teacher's grading practice bears an increasingly smaller weight on the job market's perception of grading standards. Then, similarly to the case with nominal grades, teachers find grades costless rewards to give and, therefore, tend to exploit good grades to their benefit (which is more student effort). But when every teacher does so: "the problem of the commons" of good grades arises, which leads to the deterioration of grading standards and grade inflation in the end.

The remainder of the paper is organized as follows. After a note on related literature about modeling grading rules, in Section 2 we present a modeling framework. In Section 3, we solve the model for the case when students value grades nominally, and in Section 4—for the case when students value grades relatively. In Section 5, we discuss the main findings of the model presented and relate them to the grading patterns observed in practice; there, we also discuss policy applications of the model. In Section 6, we review existing empirical evidence on mismatch between students' grades and their abilities. The last section concludes the study.

[4] Dickson (1984). For more on grade inflation, see, e.g., Sabot & Wakeman-Linn (1991) or Kuh & Hu (1999) and other references cited therein. The idea that more student effort could be behind higher average grades is also rejected by Babcock & Marks (2010), where they show that over the period 1961–2003 time spent for study by full-time students in the US gradually decreased from 40 to 27 hours.

[5] ... which is consistent with the observation of declining college entrance test scores (Wilson (1999)) and the steadily increasing number of students enrolled in US degree-granting institutions both in absolute and relative (as a percentage of high school graduates) numbers: see *Table 200. Recent high school completers and their enrollment in college, by sex: 1960 through 2008* of 2009 Digest of Education Statistics, http://nces.ed.gov/programs/digest/d09/tables/dt09_188.asp.

[6] ... as documented in Babcock & Marks (2010).

[7] The number of degree-granting institutions in the US gradually increased from 563 in 1869–1870 to 4352 in 2007–2008, see *Table 188. Historical summary of faculty, students, degrees, and finances in degree-granting institutions: Selected years, 1869-70 through 2007-08* of 2009 Digest of Education Statistics, http://nces.ed.gov/programs/digest/d09/tables/dt09_188.asp.

Literature

The modeling approach of a teacher-student relationship undertaken in this paper is novel in economics literature, but it does not stand alone in taking the perspective of a principal-agent model to study the question of optimal grading rules.[8] Dubey & Geanakoplos (2010) target the question what grading scale (finer or coarser) a teacher should use in order to induce a higher effort from her students. They approach the problem of optimal grading schemes from a perspective different from ours: they model a teacher-student relationship as a game of status with stochastic output similar to a tournament. In their multiple-agent model, a student's utility of a grade depends on his or her class ranking, i.e., status, resulting from the grade rewarded (but not on the grade *per se* even if it carries the same ability signal irrespective of the distribution of grades in the class). In addition, in their model the teacher aims to incentivize all her students to put in maximal effort rather than to obtain the highest expected effort. Given these modeling differences, we draw different conclusions about optimal grading schemes. Dubey & Geanakoplos (2010) find that teachers should use coarse grading schemes and "pyramid" the allocation of grades: in equilibrium the highest grade would be available to fewer students than the second-highest grade, and so on.[9] Our model predicts that teachers should apply coarse grading schemes, but only when they can "costlessly" reward the student (e.g., when an individual teacher's grading practice cannot affect the perception of the job market about the value of grades), otherwise they should apply a fine grading rule. In addition, in the case with nominal grades, we do not find "pyramiding" to be an optimal grading rule, especially, when there is a large mass of less able students in the class.

Another strand of literature considers grades as cheap talk between university administration and the job market about student abilities (Chan *et al.* (2007), Ostrovsky & Schwarz (2010)). The objective of university administration is to achieve higher average job placements for their students. As shown in Ostrovsky & Schwarz (2010), university administration may find it beneficial to compress information sent to the job market by coarsening grading schemes. Chan *et al.* (2007) shows how competition among schools can further worsen the problem of compressed grades resulting in the grade inflation. Even though in our paper the cheap-talk element of grades is not present: a teacher's and the job market's interests are orthogonal to one another; it could be, however, an interesting extension to our model, as discussed later in the text.

[8] However, it needs to be reckoned that not much theoretical work has been done on modeling a teacher-student relationship as a principal-agent model on its own: it has rather been modeled as part of a larger setting involving potential employers or university administration, see McKenzie & Tullock (1981, Ch. 17) for general discussion or recent papers Chan *et al.* (2007) and Ostrovsky & Schwarz (2010), discussed below.

[9] Moldovanu *et al.* (2007), a study on contests for status rewards, makes a similar prediction as well.

2 Framework

There is a teacher teaching her class of students a particular subject. The teacher's goal is to pass on knowledge in her subject to the students. She has a technology—the exam test—that allows her to assess the knowledge level attained by a student from his test score $x \in [0, \overline{x}]$. (The upper bound on test scores, $\overline{x}$, is large enough to allow for an interior solution.) We assume that the teacher's technology is perfect in the sense that there is a deterministic and direct relationship between test scores and knowledge levels, both of which, therefore, are used synonymously throughout. Achieving a test score x comes to a student at an effort cost of $C(x, \theta)$, where parameter θ is the student's privately known ability level for the subject studied, distributed in the population according to a common prior distribution F over the ability space $\Theta = \left[\underline{\theta}, \overline{\theta}\right]$, $\overline{\theta} > \underline{\theta} > 0$. The properties of the effort cost function C are $C_x > 0, C_{xx} > 0, C_\theta < 0$, and $C_{x\theta} < 0$.

Suppose that students select the teacher's class for exogenous reasons (e.g., the class is compulsory in their curriculum).[10] The reward pursued by every student is the teacher's assessment—grade $r \in [0, \overline{r}]$—of his class performance, i.e., knowledge attained. The upper bound on grades, $\overline{r}$, is assumed to be institutionally preset, as is the very grading framework: the assessment of student performance needs to done in the form of grades only.

As already said in the introduction, we separately consider two sources of grade value for students: 1) a grade is of a value on its own (Section 3. Nominal Value of Grades) and 2) the value of a grade comes in the form of an ability signal inferred by the job market (Section 4. Relative Value of Grades). The exact forms of the student utility functions of grades are given in the corresponding sections.

The teacher and students' relationship develops as follows. First, the teacher sets up a grading rule that assigns grades r to test scores x. Then, after observing the grading rule, each student decides on a learning effort $C(x, \theta)$ to achieve the test score x rewarding the grade r. The teacher's objective is to maximize her expected utility, which increases in students' knowledge levels—equivalently, in their test scores—and every student's objective is to maximize his utility of a grade less the effort cost spent to obtain it.

Here, we impose some further structure on the model. The cumulative ability distribution function F is twice differentiable and its probability density function f is strictly positive everywhere ($f > 0$). In addition, we impose the assumption that the hazard rate, $h(\theta) = f(\theta)/(1 - F(\theta))$, monotonically increases in ability, i.e., $h'(\theta) \geq 0$. The student

[10] Arcidiacono (2004) argues that students select their majors out of intrinsic preference for them, for which he provides empirical evidence showing that expected future earnings do not explain students' school and major choices. At the same time, Sabot & Wakeman-Linn (1991) empirically show that students' choice of classes *not required for their majors* is responsive to their grade expectations, which we do not model here essentially restricting attention to the choice of classes in students' major.

effort cost function C is separable in test score x and ability θ and takes the form of

$$C(x, \theta) = \frac{y(x)}{\theta},$$

where y is an increasing and strictly convex function of x. Then, denote the teacher's utility of test score x elicited from a student by function $V : [0, \overline{x}] \rightarrow \mathbb{R}^+$, $V_x > 0$, and assume, until further notice, that it is linear in x :

$$V(x) = x. \tag{1}$$

The teacher's linear utility can be interpreted that she equally cares about low- and high-ability students' performance. Later in the text, when discussing policy applications of the model, the linear case serves as a benchmark for the cases when 1) function V is convex (the teacher puts more weight on high-ability students), and 2) function V is concave (the teacher puts more weight on low-ability students).

3 Nominal Value of Grades

In this section, we model that students value grades at their face value and derive a higher utility from a higher grade independently of the grading rule applied by the teacher. As already discussed in the introduction, this assumption allows us to treat the teacher's problem of designing grading rules as a *single-student* agency problem.

Next, when formulating the teacher's utility maximization problem—designing the student-knowledge-maximizing grading rule—we, without loss of generality, restrict the set of grading rules to direct grading rules $\{\mathbf{x}, \mathbf{r}\}$, where test score schedule $\mathbf{x} : \Theta \rightarrow [0, \overline{x}]$ and grade schedule $\mathbf{r} : \Theta \rightarrow [0, \overline{r}]$ belong to the class of piecewise continuously differentiable functions, denoted by $\mathcal{C}^1_{\mathbf{x}}$ and $\mathcal{C}^1_{\mathbf{r}}$, respectively. Furthermore, from the Revelation Principle we know that an optimal grading rule $\{\mathbf{x}, \mathbf{r}\}$ needs to impose on the student the truthful revelation of own ability type.

Given a grading rule $\{\mathbf{x}, \mathbf{r}\}$, the student's net utility of reporting a type $\hat{\theta}$ is equal to

$$U(\theta, \hat{\theta}) = \mathbf{r}(\hat{\theta}) - \frac{y(\mathbf{x}(\hat{\theta}))}{\theta}, \tag{2}$$

where parameter θ is his true ability type. The student's reservation utility of participating in the class is normalized to zero (for all ability types).

3.1 The Teacher's Problem

The teacher maximizes with respect to a grading rule $\{\mathbf{x}, \mathbf{r}\} \in \mathcal{C}^1_{\mathbf{x}} \times \mathcal{C}^1_{\mathbf{r}}$ her expected utility

$$\int_{\underline{\theta}}^{\overline{\theta}} \mathbf{x}(\theta) dF(\theta) \tag{3}$$

subject to

$$U(\theta, \theta) \geq U(\theta, \hat{\theta}), \tag{4}$$

$$U(\theta, \theta) \geq 0, \tag{5}$$

$$0 < \mathbf{r}(\theta) \leq \overline{r}, \text{ for all } \theta \text{ and } \hat{\theta} \text{ in } \Theta. \tag{6}$$

In the above, (4) is the student's incentive compatibility constraint, (5) — individual rationality constraint, where the utility function U is defined in (2); the last constraint imposes an upper bound on the teacher's rewards (though already imposed by requiring $\mathbf{r} \in \mathcal{C}^1_{\mathbf{r}}$).

Under this specification, the model resembles a standard static principal-agent model with hidden information (monopolistic screening problem) except for the transfer structure. The distinct feature of this model is that, unlike in most agency models, the transfer function $\mathbf{r}$ does not enter the principal's (teacher's) utility function but only the agent's (student's), meaning that the principal does not "pay" to motivate the agent. In other words, rewards are costless for the principal to give but of a value to the agent.

Therefore, unlike in models with monetary transfers, in this model with costless rewards we do not have the intercomparison of the agent's and principal's utilities. To solve the model, we approach it differently from the standard solution method attributable to Mirrlees (1971), which main idea is to obtain a functional equation with one unknown by merging the agent's and principal's optimization problems through the transfer function. The key element in solving our model is the observation that it must be optimal for the principal to reward the highest grade of $\overline{r}$ to the most efficient agent, i.e., in the solution we have $\mathbf{r}(\overline{\theta}) = \overline{r}$ because the reward is costless. This observation allows us to reduce all the constraints (4)–(6) into a single constraint and solve the model using the standard Langragean methods.

The proposed refinement that, unlike the agent, the principal is indifferent to a transfer between them is by no means new in the contract theory literature. It was studied in Guesnerie & Laffont (1984), which provides an all-encompassing solution to a broadly defined principal-agent problem. In particular, they distinguish between "type A" and "type B" preferences, where with the former preferences the principal's utility does not depend on a transfer, while with the latter (conventional) preferences it does. In their study, however, the "type A" preferences are primarily used to analyze a social planner's problem of social welfare maximization. There, a transfer between the social planner (principal) and the agent is equivalent, figuratively speaking, to distributing money between two pockets of the same jacket, leaving the social welfare intact. Therefore, the framework of Guesnerie & Laffont (1984) does not apply to the problem studied here. In our model, the principal is, in fact, more of "type B", i.e., she cares only about her own utility but does not pay for motivating the agent.

Furthermore, the problem stated here closely connotes with the textbook problem of pricing a single indivisible object, to which a seller attaches no value (see, e.g., Krishna (2002)). The seller's objective is to maximize her expected revenue received from selling the object to a single buyer. Then, in the language of the price-setting framework, in our problem (3)–(6) the function $\mathbf{x}$ reads like a payment schedule, $\mathbf{r}$ — a probability schedule of selling the object (with the upper bound set at $\overline{r} = 1$, respectively). An important difference between our model and the pricing model in the way it is normally approached is that in our model a score allocation x enters the agent's utility function in a nonlinear way, whereas in the other model it enters linearly. As such, if applied to the pricing-setting problem, with some adjustments of the agent's utility function our model, as we are going to see next, would render that the seller can offer the object to some agent types with a positive probability below 1, unlike in the pricing model with linear utilities.

3.2 Solution

We present the solution, the grading rule $\{\mathbf{x}, \mathbf{r}\}$ maximizing (3) subject to (4)–(6), in Proposition 1 below, relegating the details of solving the model to the Appendix. The main property of the solution is the pooling of ability types from a non-empty interval $\left[\theta^*, \overline{\theta}\right]$ for the highest score-grade allocation, which we discuss more thoroughly later.

Proposition 1 *The score-grade allocations* $(\mathbf{x}(\theta), \mathbf{r}(\theta))$*, solving the optimization problem (3)–(6), are characterized*

- *for ability types* θ *in* $[\underline{\theta}, \theta^*)$*, where*

$$\theta^* = \min\{\theta : (1 - F(\theta))/(\theta f(\theta)) \leq 1, \theta \in \left[\underline{\theta}, \overline{\theta}\right]\}, \tag{7}$$

  by

$$\mathbf{x}(\theta) = y_x^{-1}\left(\frac{f(\theta)\theta^2 y_x(\mathbf{x}(\theta^*))}{f(\theta^*)\theta^{*2}}\right) \tag{8}$$

  for the score allocations $\mathbf{x}$*, and by*

$$\mathbf{r}(\theta) = \frac{y(\mathbf{x}(\theta))}{\theta} + \int_{\underline{\theta}}^{\theta} \frac{y(\mathbf{x}(\tilde{\theta}))}{\tilde{\theta}^2} d\tilde{\theta} \tag{9}$$

  for the grade allocations $\mathbf{r}$*; and*

- *for ability types θ in $\left[\theta^*, \overline{\theta}\right]$ by the grade $\mathbf{r}(\theta) = \overline{r}$ and the score allocation $\mathbf{x}(\theta) = \mathbf{x}(\theta^*)$, where $\mathbf{x}(\theta^*)$ is found from*

$$\overline{r} - \frac{y(\mathbf{x}(\theta^*))}{\theta^{*2}} - \int_{\underline{\theta}}^{\theta^*} \frac{y(\mathbf{x}(\theta))}{\theta^2} d\theta = 0. \tag{10}$$

Proof. See Appendix A.[11] ■

Efficiency

In this model with complete information, the first-best grading rule is determined by the student's participation constraints only. Since the rewards are costless and the teacher's utility increases in score x, the teacher would set up a grading rule that gives the highest grade of $\overline{r}$ to every student type against the effort level determined by the corresponding participation constraint. From the solution presented in Proposition 1, we can immediately see that the efficient score levels can only be achieved for the lowest type in the event when the teacher pools all the types, i.e., θ^* in (7) is equal to $\underline{\theta}$. In particular, the "no distortion at the top" property, observed in the standard model, never holds here.

3.3 Results

Below we summarize the main properties of the optimal grading rule with nominal grades, presented in Proposition 1.

Pooling at the top

To make it general, one of the main properties of the model studied is the optimality of a uniform allocation among most efficient agent types when the principal does not bear or does not internalize the cost of rewarding the agent. This result is in contrast to the "no distortion at the top" property of optimal contracts obtained in most agency models with costly rewards and hidden information.[12]

Result 1 *In a single-agent and hidden-information agency problem with costless rewards and a continuum of types, the principal pools some of the most efficient agent types for a uniform allocation.*

[11] For convenience, we assume that the teacher always finds it optimal to offer non-zero allocations to all the student types. Also, the problem is solved for the case with the linear utility function V. However, this has no effect on the qualitative properties of the solution obtained, which are invariant to the form of the teacher's utility function V (which needs to be "less convex" than the effort cost function C). In particular, the starting point of the pooling interval θ^* remains the same for any functional form of V.

[12] While the "no distortion at the top" property is characteristic of principal-agent models with monetary rewards, see, e.g., Mirrlees (1971), it has also been shown to hold for agency problems with status incentives, see Moldovanu *et al.* (2007). But recently, there have also been papers in which this property does not hold in the optimum, see Levin (2003) or MacLeod (2003), where the result hinges on the assumption that the agent's effort is not verifiable unlike in our model.

The proof of this result has been given when deriving the condition for the starting point of the pooling interval (7) in Proposition 1 (see Appendix A), where we show that a uniform score-grade allocation $(\mathbf{x}(\theta^*), \overline{r})$ applies to all the agent types from the non-empty interval $[\theta^*, \overline{\theta}]$ (the starting point θ^* is bound to be strictly less than the highest ability type $\overline{\theta}$, since $(1 - F(\overline{\theta}))/(\overline{\theta} f(\overline{\theta})) = 0 < 1$). Moreover, neither the existence of an upper bound on rewards nor its size, as imposed by constraint (6), is central to the result, what is crucial is the costlessness of rewards. (In a principal-agent model with costly rewards, imposing an upper bound on a reward function does not lead to pooling among most efficient agent types as long as this constraint is not binding, i.e., when the upper bound is large enough.)

The finding that there is no perfect screening among the most efficient agent types should not be surprising. Suppose it were the case that in the solution only the most efficient type received the highest reward. To make this allocation incentive compatible, the principal would need to suppress the motivation of other types in order to refrain the most efficient type from misreporting. But, as an alternative to the perfect screening, consider the principal marginally "tilting up" the schedule of all the allocations but the last one—done at no cost—against the corresponding decrease in the performance allocation of the most efficient agent type. This change in the allocations is sure to be expected performance increasing because the gain from it—the increases in performance levels for almost all types—outweighs its corresponding loss—the decrease in performance level of the most efficient type happening with a very small probability. As a result, the principal finds it optimal to increase the probability mass of agent types subject to the highest reward until the gains and losses described offset each other.

Referring back to the teacher-student relationship, where lenient grading is a widespread phenomenon, pooling most efficient types is even more prevalent when the distribution for abilities is more skewed to the end of low types as discussed next.

Mismatch between grades and abilities

Here, we establish a relationship between the optimal grading rule and student ability distribution, which has a strong empirical support from the literature on educational measurement, described later in the paper.

Consider two classes of students, who come from two different student populations, where abilities are distributed on the same support Θ according to distributions F_1 and F_2, respectively. Denote the student types from the two classes by θ_1 and θ_2, respectively, and let the student type θ_2 be smaller than θ_1 in the likelihood ratio order, i.e.,

$$\frac{f_2(\theta)}{f_1(\theta)} \text{ decreases for all } \theta \text{ in } \Theta,$$

where f_1 and f_2 are the probability density functions of the corresponding distributions.

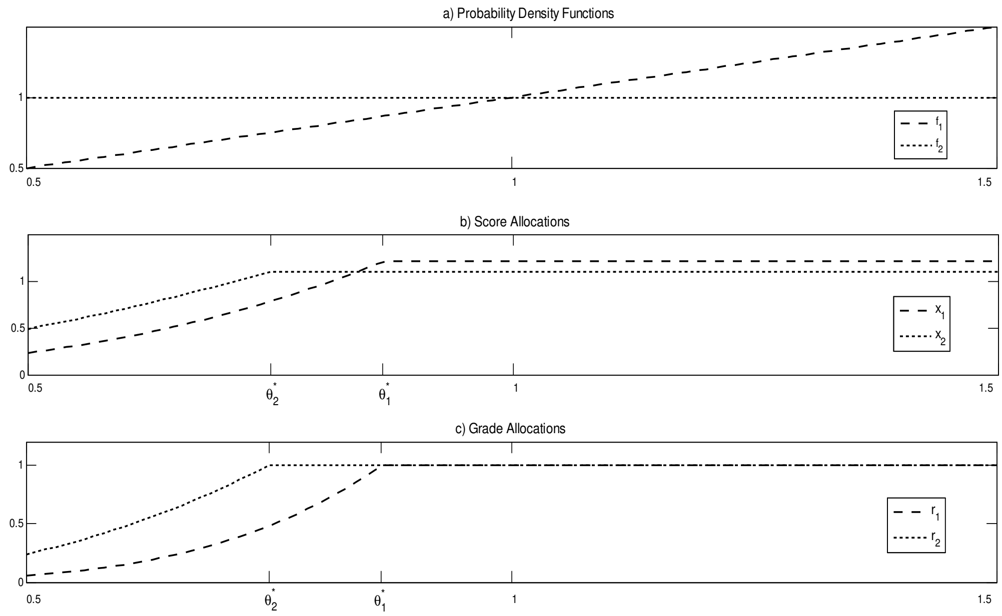

Figure 1: Optimal Grading Rules

The interpretation of this stochastic dominance condition is that students from the first class are held more able than those from the second class. (Formally, this condition implies that $\int_{\theta'}^{\theta''} \theta f_1(\theta) d\theta \geq \int_{\theta'}^{\theta''} \theta f_2(\theta) d\theta$ for any interval $[\theta', \theta''] \subseteq \Theta$; in words, for any restriction of the ability space the expected student ability in the first class is greater than that in the second class.)

Let $\{\mathbf{x}_1, \mathbf{r}_1\}$ and $\{\mathbf{x}_2, \mathbf{r}_2\}$ be the solutions to the optimization problem (3)–(6) for the two classes, respectively. Then, the following holds.

Result 2 *If the student type θ_2 is smaller than the student type θ_1 in the likelihood ratio order, then the optimal grade allocations in the two classes satisfy $\mathbf{r}_2(\theta) \geq \mathbf{r}_1(\theta)$ for every student type θ in Θ.*

Proof. See Appendix B. ■

To put it in words, this result says that the lower the expectations the teacher holds about her student abilities, the more lenient she should be when grading. The intuition behind the optimality of more lenient grading rules in less able classes comes from the teacher's attempt to extract more effort from more numerous lower-ability student types, and *vice versa*.

Results 1 and 2 are illustrated in Figure 1. It depicts the optimal grading rules for two classes with student ability type space $\Theta = [0.5, 1.5]$; the student ability in the first class, θ_1, is distributed according to the distribution with $f_1(\theta) = \theta$, $\theta \in \Theta$, and in the second class, the student ability θ_2 is distributed uniformly over Θ (the top diagram, the dashed line for the first class and the dotted line for the second class). The effort cost function takes the form of $C(x, \theta) = x^2/(2\theta)$ and the upper bound on grades is set to $\overline{r} = 1$. The middle diagram of Figure 1 shows the optimal score allocations, and the bottom diagram shows the optimal grade allocations for the two classes, respectively. As we can see, both teachers pool some of the most efficient student types. The teacher of the first class, however, offers the highest grade of $\overline{r}$ to fewer student types but against a higher performance level, while the teacher of the second class optimally chooses to be more lenient. Unlike the teacher of the more able class, the second teacher also offers a flatter score-grade schedule in her attempt to extract more effort from less able but relatively more numerous student types.

4 Relative Grades

In this section, we study the situation when students value grades only as ability signals to the job market. Essentially, we close the model of Section 3 by introducing the job market, which defines the value of a grade. In what follows, we distinguish between two cases based on the scope of the job market's knowledge: 1) observable grading rules—the job market observes the exact grading rule applied to grade students—and 2) unobservable grading rules—the job market does not observe the grading rule applied and infers a student's expected ability from his grade based on its own perception of grading standards used by teachers.

4.1 Observable Grading Rules

Here, we modify the previous model by having that 1) the student values grades as ability signals, and 2) the job market observes the grading rule set up by the teacher and correctly infers the student's (expected) ability from his grade, as defined below. Therefore, unlike in the previous section, here the value of a grade is dependent on the distribution of grades across different ability types or, in other words, on the stringency of the grading rule applied by the teacher.

Suppose the teacher designs an incentive compatible grading rule $\{\mathbf{x}, \mathbf{r}\}$. Let $\mathcal{R}(\mathbf{r})$ denote the range of the grade schedule $\mathbf{r}$. If $r \in \mathcal{R}(\mathbf{r})$, $\mathbf{r}^{-1}(r)$ is the set of all $\theta \in \Theta$ such that $\mathbf{r}(\theta) = r$. Let $\theta_*(r) = \inf(\mathbf{r}^{-1}(r))$ and $\theta^*(r) = \sup(\mathbf{r}^{-1}(r))$ denote the lowest and highest types (at the limit) subject to the same grade r. Using the monotonicity of the grade schedule $\mathbf{r}$, we define the ability type $\theta^{\mathbf{r}} : [0, \overline{r}] \to \Theta$, inferred by the job market from a grade r, by

$$\theta^{\mathbf{r}}(r) = \begin{cases} \mathbf{r}^{-1}(r) & \text{if} \quad r \in \mathcal{R}(\mathbf{r}) \text{ and } \theta_*(r) = \theta^*(r); \\ \frac{\int_{\theta_*(r)}^{\theta^*(r)} \theta f(\theta) d(\theta)}{F(\theta^*(r)) - F(\theta_*(r))} & \text{if} \quad r \in \mathcal{R}(\mathbf{r}) \text{ and } \theta_*(r) \neq \theta^*(r); \\ 0 & \text{if} \quad r \notin \mathcal{R}(\mathbf{r}). \end{cases} \tag{11}$$

As a technical detail, we set $\theta^{\mathbf{r}}(r) = 0$ if $r \notin \mathcal{R}(\mathbf{r})$ to mean that if a student's grade is outside the range of the grading rule applied then this grade as a signal is meaningless (i.e., the job market infers that the student has no ability in the subject in question).

The above definition also defines a θ-type student's utility of a grade given a grading rule $\{\mathbf{x}, \mathbf{r}\}$, and his net utility of reporting a type $\hat{\theta}$ is equal to

$$U^{\mathbf{r}}(\theta, \hat{\theta}) = \theta^{\mathbf{r}}(\mathbf{r}(\hat{\theta})) - C(\mathbf{x}(\hat{\theta}), \theta). \tag{12}$$

The teacher's problem

As before, the teacher's problem is to set up a grading rule that elicits from the student the highest expected performance on the test, i.e., the teacher maximizes her expected utility with respect to a grading rule $\{\mathbf{x}, \mathbf{r}\} \in \mathcal{C}^1_{\mathbf{x}} \times \mathcal{C}^1_{\mathbf{r}}$

$$\int_{\underline{\theta}}^{\overline{\theta}} \mathbf{x}(\theta) dF(\theta) \tag{13}$$

subject to

$$U^{\mathbf{r}}(\theta, \theta) \geq U^{\mathbf{r}}(\theta, \hat{\theta}), \tag{14}$$

$$U^{\mathbf{r}}(\theta, \theta) \geq 0, \text{ for all } \theta \text{ and } \hat{\theta} \text{ in } \Theta. \tag{15}$$

In the above, the constraints are the student's incentive compatibility and individual rationality constraints, respectively, where the student's utility function $U^{\mathbf{r}}$ is defined in (12).

In this version of the model, we can simplify the teacher's problem (13)–(15) by making it be maximized with respect to a score schedule $\mathbf{x}$ only. The reason for this is that a grade does not bear any value on its own and, as a result, a score schedule $\mathbf{x}$ can be implemented by any grade schedule $\mathbf{r}$ isomorphic to $\mathbf{x}$ achieving the same utility levels. Therefore, without loss of generality, in the teacher's problem we restrict the set of grading rules $\mathcal{C}^1_{\mathbf{x}} \times \mathcal{C}^1_{\mathbf{r}}$ to such grading rules $\{\mathbf{x}, \mathbf{r}\}$, where the grade schedule $\mathbf{r}$ takes the form of

$$\mathbf{r}(\theta) = \overline{r} \frac{\mathbf{x}(\theta)}{\mathbf{x}(\overline{\theta})}, \text{ for any } \theta \in \Theta. \tag{16}$$

All in all, the teacher maximizes (13) with respect to $\mathbf{x} \in \mathcal{C}^1_{\mathbf{x}}$ subject to (14) and (15) with the grade schedule $\mathbf{r}$ imposed by (16).

Solution

If the student values a grade for its ability signal, the teacher, when designing a grading rule, needs to take into account the effect the grading rule itself has on the value of a grade for the student as determined by (11). Unlike in the previous model with nominal grades, it follows that the more lenient the grading rule the teacher sets up, the lower the utility the student gets from a given grade, adversely affecting his learning effort choice decision. Below, we give the condition when this downside effect of leniency in grading—the job market's "degrading" of students' grades—outweighs the upside effect—more effort extraction from lower ability types.

First, we solve for the optimal score schedule $\mathbf{x}$ under the conjecture that the teacher perfectly screens all the types.

Conjecture 1 *The grading rule $\{\mathbf{x}, \mathbf{r}\}$ solving the teacher's problem (13)–(16) screens every ability type θ distinctly.*

Among strictly monotone score functions $\mathbf{x} \in \mathcal{C}^1_{\mathbf{x}}$, the solution to the teacher's problem is uniquely characterized by the student's constraints (14) and (15) only. In the solution $\mathbf{x}$, the utility to the student of ability θ from reporting being a type $\hat{\theta}$ is given by

$$U^{\mathbf{r}}(\theta, \hat{\theta}) = \hat{\theta} - \frac{y(\mathbf{x}(\hat{\theta}))}{\theta},$$

and, at a point of differentiability, the incentive compatibility constraint implies that

$$\frac{\partial}{\partial \hat{\theta}} U^{\mathbf{r}}(\theta, \theta) = 0,$$

or

$$y_x(\mathbf{x}(\theta))\mathbf{x}_\theta(\theta) = \theta.$$

Next, it has to be that the individual rationality constraint of the lowest ability type $\underline{\theta}$ needs to be binding (or, more precisely, that of the least efficient type "contracted upon" by the teacher, which we assume to be $\underline{\theta}$). If there are no discontinuities—which we check if it is the case later—the solution to the above differential equation together with the binding individual rationality constraint is

$$\mathbf{x}(\theta) = y^{-1}\left(\frac{\theta^2 + \underline{\theta}^2}{2}\right), \tag{17}$$

for every θ in Θ. Since the derivative of $\mathbf{x}$ in (17) is positive for every type θ, the condition that $\mathbf{x}$ be strictly monotone holds.

But suppose that there are (simple) discontinuities in the score schedule $\mathbf{x}$ that solve the teacher's problem. Denote the discontinuity point closest to the type $\underline{\theta}$ by θ', and let $\mathbf{x}(\theta'-) \neq \mathbf{x}(\theta')$, where $\mathbf{x}(\theta'-)$ is the left-hand limit (the subsequent argument with straightforward alterations also holds for the case $\mathbf{x}(\theta'+) = \mathbf{x}(\theta')$, where $\mathbf{x}(\theta'+)$ is the right-hand limit). Then, (17) holds only for types θ in $[\underline{\theta}, \theta')$. But the allocation $\mathbf{x}(\theta')$ cannot be greater than the allocation determined by (17) for $\theta = \theta'$, because $\mathbf{x}(\theta')$ would not be incentive compatible: we could find a type $\tilde{\theta} \in [\underline{\theta}, \theta')$ such that $U^{\mathbf{r}}(\theta', \hat{\theta}) > U^{\mathbf{r}}(\theta', \theta')$

for every $\hat{\theta} \in (\tilde{\theta}, \theta')$. Then, $\mathbf{x}(\theta')$ should be smaller than the allocation determined by (17) for $\theta = \theta'$, which, though, is suboptimal from the teacher's perspective. Hence, there cannot exist a discontinuity at θ', nor at any other point for the same reason.

Therefore, under the conjecture that all types are screened perfectly, (17) uniquely characterizes the optimal score schedule $\mathbf{x}$ for every θ in Θ. Next, we need to give a condition when this conjecture is valid. It turns out that the sufficient condition is the convexity of the score schedule $\mathbf{x}$ in (17).

Proposition 2 *If the score schedule* $\mathbf{x}$ *defined in (17) is convex, then it is the solution to the teacher's problem (13)–(15) with the grade schedule* $\mathbf{r}$ *imposed by (16).*

Proof. See Appendix C. ■

The idea of the proof is that, once the convexity condition is met, a grading rule containing a uniform allocation for some types can be improved upon by separating those types with distinct (and incentive-compatible) allocations as in (17). Furthermore, the convexity condition is equivalent to requiring that the marginal disutility from effort under $\mathbf{x}$ as in (17) decrease in ability

$$\frac{\partial}{\partial \theta}(C_x(\mathbf{x}(\theta), \theta)) < 0 \tag{18}$$

or, under the functional assumption of $C(x, \theta) = y(x)/\theta$ and dropping the arguments,

$$y_{xx} \leq \left(\frac{y_x}{\theta}\right)^2,$$

which is to require that the effort cost function be not "too convex" in scores x. (For instance, this condition is met for the effort cost function quadratic in x and, correspondingly, for other cost functions "less convex" than the quadratic one).

However, if the score schedule $\mathbf{x}$ defined in (17) is not convex everywhere, then at the restrictions of the type space Θ where it is concave, we obtain pooling of types (by reversing the argument of the proof of Proposition 2). Moreover, with a general form of the teacher's utility function V in (1) the condition for screening types is that the function

$$V\left[y^{-1}\left(\frac{\theta^2 + \underline{\theta}^2}{2}\right)\right]$$

is convex in θ. With a concave (convex) utility function V, this condition is less (more) likely to hold.

Given that the convexity condition holds, the finding that the teacher screens all the student types—when the student values grades for their signaling value and the job market can observe the grading rule designed—is in stark contrast to the results obtained for the case with nominal grades, where pooling of most efficient types is always optimal (see Result 1 of the previous section). This difference in the optimal grading rules results from the difference in the utility of a grade perceived by the student in the two models studied. However, as we show next, if the job market cannot observe grading rules designed by individual teachers and, correspondingly, relate a student's grade to the grading rule applied to grade him, then we again observe coarse grading rules with pooling types even if the student values a grade for its signaling value only.

4.2 Unobservable Grading Rules

Here, we study the situation when the job market does not observe the grading rule that the teacher has designed to grade her student. (This can happen, for instance, when there are too many teachers for the job market to distinguish among.) Instead, we assume that the job market holds its own perception of the grading standard applied, which we denote by $\boldsymbol{\rho} \in \mathcal{C}^1_{\mathbf{r}}$, and the student's expected ability it infers from a grade r is, accordingly, $\theta^{\boldsymbol{\rho}}(r)$ defined by (11) for the grading standard $\boldsymbol{\rho}$ (i.e., in definition [11] $\mathbf{r}$ is replaced with $\boldsymbol{\rho}$). The ability inferred, $\theta^{\boldsymbol{\rho}}(r)$, also defines the student's value of a grade r. The assumption is that the grading standard $\boldsymbol{\rho}$ is public information.

Now, the teacher's problem is to design a grading rule $\{\mathbf{x}, \mathbf{r}\} \in \mathcal{C}^1_{\mathbf{x}} \times \mathcal{C}^1_{\mathbf{r}}$ such that it maximizes her expected utility against the grading standard $\boldsymbol{\rho}$

$$\int_{\underline{\theta}}^{\overline{\theta}} \mathbf{x}(\theta) dF(\theta) \tag{19}$$

subject to

$$\begin{aligned} \theta^{\boldsymbol{\rho}}(\mathbf{r}(\theta)) - C(\mathbf{x}(\theta), \theta) &\geq \theta^{\boldsymbol{\rho}}(\mathbf{r}(\hat{\theta})) - C(\mathbf{x}(\hat{\theta}), \theta), &(20)\\ \theta^{\boldsymbol{\rho}}(\mathbf{r}(\theta)) - C(\mathbf{x}(\theta), \theta) &\geq 0, \text{ for all } \theta \text{ and } \hat{\theta} \text{ in } \Theta. &(21) \end{aligned}$$

Unlike in the previous problem (13)–(15), here the teacher does not internalize the effect her grading rule has on the student's utility of a grade, $\theta^{\boldsymbol{\rho}}$, which is determined by the job market's grading standard $\boldsymbol{\rho}$. Therefore, the teacher finds grades costless to reward. In this respect, the problem turns similar to the one with nominal grades (3)–(6), and, therefore, the solution method, applied in Appendix A when solving the model with nominal grades, applies here, too.

In what follows, we are interested in the consistent grading standard $\boldsymbol{\rho}$, which needs to be equal to the grading schedule $\mathbf{r}$ solving the teacher's problem (19)–(21) against the grading standard $\boldsymbol{\rho}$.

Consistent grading standard

We make the following definitions.

Definition 1 *Let a mapping $\pi : \mathcal{C}^1_{\mathbf{r}} \to \mathcal{C}^1_{\mathbf{r}}$ map a grading standard $\boldsymbol{\rho}$ into the grade schedule $\mathbf{r}$ of the grading rule solving the teacher's problem (19)–(21) against the grading standard $\boldsymbol{\rho}$.*

For analytical convenience, we assume that for any $\boldsymbol{\rho}$ there is a unique solution to the teacher's problem and, accordingly, a unique grade schedule $\mathbf{r}$.

Definition 2 *A grading standard $\boldsymbol{\rho}$ is consistent if it is a fixed point of the mapping π:*

$$\boldsymbol{\rho} = \pi(\boldsymbol{\rho}).$$

In words, a grading standard $\boldsymbol{\rho}$ is consistent if the grade schedule $\mathbf{r}$ of the teacher's optimal grading rule against the grading standard $\boldsymbol{\rho}$ is equal to the grading standard $\boldsymbol{\rho}$ itself.

We can immediately establish the following property of a consistent grading standard.

Result 3 *A grading standard ρ with perfect screening of ability types cannot be consistent.*

This result is a direct consequence of Proposition 1 (and Result 1), where we show that in a model with costless rewards pooling of types is inevitable. In the event that the job market perceives a grading standard $\boldsymbol{\rho}$ that all student types are screened perfectly, i.e., $\boldsymbol{\rho}(\theta_1) > \boldsymbol{\rho}(\theta_2)$ if $\theta_1 > \theta_2$, $\theta_1, \theta_2 \in \Theta$, the teacher's problem (19)–(21) becomes similar to that with nominal grades (3)–(6), where the teacher can reward the student with any ability signal $\theta^{\boldsymbol{\rho}}$ by manipulating the grade schedule $\mathbf{r}$. More precisely, we can make the following identity transformations of (19)–(21) to make it look like (3)–(6): set $\boldsymbol{\theta}(\theta) \equiv \theta^{\boldsymbol{\rho}}(\mathbf{r}(\theta))$ and introduce a constraint $\underline{\theta} \leq \boldsymbol{\theta}(\theta) \leq \overline{\theta}$ for all $\theta \in \Theta$. Then, the teacher maximizes her expected utility with respect to $\{\mathbf{x}, \boldsymbol{\theta}\}$, and the optimal signal schedule $\boldsymbol{\theta}$ would pool types θ in $[\theta^*, \overline{\theta}]$, where θ^* defined by (7) in Proposition 1, for the reward of $\overline{\theta}$, implying that the grading standard $\boldsymbol{\rho}$ cannot be consistent.

Change in grading standards

In the following subsection, we establish the existence result of a consistent grading standard. Here, we attempt to analyze iteratively the dynamics of grading standards under the adaptive expectations framework, in particular, to see the direction of change of the interval of ability types pooled for the highest grade (more precisely, for the highest ability signal inferred by the job market).

Suppose the initial grading standard $\boldsymbol{\rho}_0 \in \mathcal{C}^1_{\mathbf{r}}$ is given and let it be a strictly increasing function of θ with $\boldsymbol{\rho}_0(\overline{\theta}) = \overline{r}$ (i.e., the job market initially perceives that the teacher screens every student type). The resultant grade schedule $\mathbf{r}_0$, maximizing the teacher's utility against the grading standard $\boldsymbol{\rho}_0$, is given by $\mathbf{r}_0 = \pi(\rho_0)$. Suppose a grade schedule $\mathbf{r}_{i-1}$,

$i \geq 1$, is given. We impose that the job market adjusts its perception of grading standards by adapting the existing grade schedule as its new grading standard, i.e., $\boldsymbol{\rho}_i = \mathbf{r}_{i-1}$, $i \geq 1$.

Consider the grade schedule $\mathbf{r}_0$. It has the properties described in Proposition 1: pooling types in $[\theta^0, \overline{\theta}]$, where $\theta^0 = \theta^*$ of Proposition 1, and screening the rest of types (assuming that $\theta^0 > \underline{\theta}$ and that the teacher offers non-zero allocations to all the types). Then, at iteration $i = 1$, the grading standard newly perceived by the job market is $\boldsymbol{\rho}_1 = \mathbf{r}_0$, which takes the form of

$$\boldsymbol{\rho}_1(\theta) = \begin{cases} \overline{r} & \text{if} \quad \theta \geq \theta^0, \\ \mathbf{r}_0(\theta) & \text{otherwise.} \end{cases}$$

The highest reward that the teacher can offer to the student against the grading standard $\boldsymbol{\rho}_1$ is $\theta^{\boldsymbol{\rho}_1}(\overline{r})$, where $\theta^{\boldsymbol{\rho}_1}$ is the (discontinuous) ability signal function defined by (11). Note that the highest reward $\theta^{\boldsymbol{\rho}_1}(\overline{r})$ has depreciated compared with the highest reward available under $\boldsymbol{\rho}_0$ and the range of rewards has contracted under $\boldsymbol{\rho}_1$, respectively.

Next, consider the grade schedule $\mathbf{r}_1 = \pi(\boldsymbol{\rho}_1)$ that the teacher sets up against the grading standard $\boldsymbol{\rho}_1$. Following the solution steps in Appendix A, one can show that the starting point θ^1 of the pooling interval $[\theta^1, \overline{\theta}]$ of the grade schedule $\mathbf{r}_1$, when $\theta^1 > \underline{\theta}$, is characterized by

$$\frac{1 - F(\theta^1)}{\theta^1 f(\theta^1)} = \frac{y_x(\mathbf{x}_1(\theta^1))}{y_x(\sup_{\underline{\theta}<\theta<\theta^1} \mathbf{x}_1(\theta))},$$

where the score allocation $\mathbf{x}_1(\theta^1)$ is for the types in the pooled interval $[\theta^1, \overline{\theta}]$, $\mathbf{x}_1(\theta)$ are the score allocations for types $\theta \in [\underline{\theta}, \theta^1)$, and $\sup_{\underline{\theta}<\theta<\theta^1} \mathbf{x}_1(\theta)$ is the supremum of score allocations $\mathbf{x}_1(\theta)$ for types θ to the left from θ^1. Since in the teacher's optimum we have $\sup_{\underline{\theta}<\theta<\theta^1} \mathbf{x}_1(\theta) < \mathbf{x}_1(\theta^1)$ because of the discontinuous grading standard $\boldsymbol{\rho}_1$, the right-hand side of the above expression is greater than 1 implying that the starting point $\theta^1$of the new pooling interval is smaller than θ^0.

Hence, the pooling interval further expands with the range of rewards—ability signals—contracting again. As a result, at the next iteration with $\boldsymbol{\rho}_2 = \mathbf{r}_1$ a teacher finds herself with fewer rewards at her disposal to give her students, arguably, leading to her further coarsening the grading schedule and the gradual depreciation of grading standards. However, to analyze tractably further change in pooling intervals $[\theta^i, \overline{\theta}]$, $i \geq 2$, or to see if the sequence of pooling-interval starting points $\{\theta^i\}$ is convergent requires placing more structure on the model (e.g., to have a complete metric space for grading rules and, then, argue that the mapping π is a contraction in attempt to apply the contraction principle). We take a different route to show the existence of a consistent grading standard, discussed in the following subsection, and here, based on the analysis above, we state the following result.

Result 4 *If a grading standard ρ is consistent, then it is the case that the job market* *perceives types θ in $[\theta^{**}, \overline{\theta}]$ pooled for the highest reward, where $\theta^{**} \leq \theta^* = \min\{\theta : (1 - F(\theta))/(\theta f(\theta)) \leq 1, \theta \in [\underline{\theta}, \overline{\theta}]\}$ with the strict inequality if $\theta^* > \underline{\theta}$.*

In words, in the case with unobservable grading rules a consistent grading standard exhibits even more pooling among the highest student types than does the grading schedule $\mathbf{r}$, characterized in Proposition 1 for the case with nominal grades.

Existence of consistent grading standards

Here, we show that there exists a consistent grading standard $\boldsymbol{\rho}$—a fixed point of π. Consider a step grading standard $\boldsymbol{\rho}$ such that

$$\boldsymbol{\rho}(\theta) = \begin{cases} r^* & \text{if } \theta \geq \theta^{\#}, \\ r_* & \text{otherwise,} \end{cases} \tag{22}$$

where $r^* > r_*$. It says that the job market perceives that student types greater than or equal to some $\theta^{\#} \in \Theta$ get the grade r^*, others—the grade r_*. (The argument that follows can be straightforwardly extended for any step grading standard with a finite number of steps.) Given a step grading standard $\boldsymbol{\rho}$ as in (22), the teacher becomes restricted to designing step grading rules $\{\mathbf{x}, \mathbf{r}\}$ of the form

$$\mathbf{x}(\theta) = \begin{cases} x^* & \text{if } \theta \geq \theta', \\ x_* & \text{otherwise,} \end{cases} \quad \text{and } \mathbf{r}(\theta) = \begin{cases} r^* & \text{if } \theta \geq \theta', \\ r_* & \text{otherwise,} \end{cases}$$

where the teacher's choice variables are scores x^*, x_* and threshold type θ'. The question is if there is a step grading standard $\boldsymbol{\rho}$ with a threshold type $\theta^{\#}$ such that the threshold type θ' of the step grading rule $\{\mathbf{x}, \mathbf{r}\}$ above solving (19)–(21) is equal to the threshold type $\theta^{\#}$ of the grading standard $\boldsymbol{\rho}$.

For a given threshold type θ', the optimal score levels x^* and x_* can be expressed as the continuous functions of θ' from the binding IR and IC constraints, respectively. Therefore, the teacher's expected utility in (19) can be expressed as the continuous function of the threshold types θ' and $\theta^{\#}$ only, which we denote by $W : \Theta \times (\underline{\theta}, \overline{\theta}) \to \mathbb{R}$ (the domain of $\theta^{\#}$ is set to be an open set, which later is expanded to Θ). Let

$$\tilde{\theta} = \arg \max_{\theta' \in [\underline{\theta}, \overline{\theta}]} W(\theta', \theta^{\#}) = \omega(\theta^{\#}),$$

which characterizes the teacher's threshold type $\tilde{\theta}$ of the optimal grade schedule $\mathbf{r}$ as a function of the threshold $\theta^{\#}$ of the grading standard $\boldsymbol{\rho}$. Now, the grading standard $\boldsymbol{\rho}$ is consistent if $\tilde{\theta} = \theta^{\#}$, i.e., if $\theta^{\#}$ is a fixed point of the solution function ω (we assume the uniqueness of the solution for every $\theta^{\#}$, which is though irrelevant for the following analysis).

By Berge's Maximum Theorem, the solution function $\omega : (\underline{\theta}, \overline{\theta}) \to \Theta$ is continuous (because the function W is continuous and the constraint correspondence is compact-valued and continuous: for any $\theta^{\#}$ the teacher's choice set of θ' is the whole ability space Θ). The next step is to apply Brouwer's Fixed Point Theorem, for which we need to expand the domain of ω to Θ. Define a function $\tilde{\omega} : \Theta \to \Theta$ by

$$\begin{aligned}
\tilde{\omega}(\overline{\theta}) &= \lim_{\theta \to \overline{\theta}} \omega(\theta), \\
\tilde{\omega}(\theta) &= \omega(\theta),\ \forall \theta \in (\underline{\theta}, \overline{\theta}), \\
\tilde{\omega}(\underline{\theta}) &= \lim_{\theta \to \underline{\theta}} \omega(\theta).
\end{aligned}$$

The function $\tilde{\omega}$ is continuous, its domain Θ is compact and convex, hence by Brouwer's Fixed Point Theorem it has a fixed point θ^{fp}

$$\theta^{fp} = \tilde{\omega}(\theta^{fp}),$$

which characterizes the consistent grading rule $\boldsymbol{\rho}$ with $\theta^{\#} = \theta^{fp}$. Having said that, we establish the following

Proposition 3 *There exists a consistent grading standard $\boldsymbol{\rho}$.*

4.3 Two-Type Example

Here, we illustrate the main results, obtained above, with a two-type example. Suppose that with positive probabilities p_1 and p_2, $p_1 + p_2 = 1$, the student's type θ takes values of θ_1 and θ_2, $0 < \theta_1 < \theta_2$, respectively. Let the student's effort cost function be $C(x, \theta) = x^2/(2\theta)$. The teacher looks for an incentive-compatible grading rule $\{\mathbf{x}, \mathbf{r}\} = \{(x_1, x_2), (r_1, r_2)\}$, where score $x_i \in [0, \overline{x}]$ and grade $r_i \in \{\underline{r}, \overline{r}\}$, $i = 1, 2$, in order to maximize the student's expected effort, which we denote by

$$\mathcal{V}(\mathbf{x}) = p_1 x_1 + p_2 x_2.$$

Let the student's utility of a grade r be measured by the expected ability level inferred by the job market. As before, we study two cases with observable and unobservable grading rules.

First, consider the case when the job market observes the grading rule $\{\mathbf{x}, \mathbf{r}\}$ applied to grade the student (as in Subsection 4.1). If the teacher separates the types, i.e., we have for the score-grade allocations $(x_1, r_1) \neq (x_2, r_2)$, then in the optimal separating grading rule $\{\mathbf{x}^s, \mathbf{r}^s\}$ the grades are $r_1^s = \underline{r}$ and $r_2^s = \overline{r}$ and the scores are pinned down from the individual rationality constraint of the low type and the incentive compatibility constraint of the high-type, yielding

$$x_1^s = \sqrt{2\theta_1^2} \text{ and } x_2^s = \sqrt{2\theta_1^2 + 2\theta_2(\theta_2 - \theta_1)}.$$

If the teacher pools the types for a uniform allocation, then in the optimal pooling grading rule $\{\mathbf{x}^p, \mathbf{r}^p\}$ we have $r_1^p = r_2^p = \overline{r}$ and

$$x_1^p = x_2^p = \sqrt{2\theta_1\hat{\theta}},$$

where $\hat{\theta} = p_1\theta_1 + p_2\theta_2$ is the ability signal inferred by the job market under pooling.

The teacher separates if the expected score under $\{\mathbf{x}^s, \mathbf{r}^s\}$ is larger than (or equal to) that under $\{\mathbf{x}^p, \mathbf{r}^p\}$, i.e., if $\mathcal{V}(\mathbf{x}^s) \geq \mathcal{V}(\mathbf{x}^p)$. To see that it is indeed the case, define a function $g : \mathbb{R}_+ \to \mathbb{R}_+$

$$g(a) = \sqrt{2\theta_1^2 + 2a(a - \theta_1)},$$

where $g(\theta_1) = x_1^s$ and $g(\theta_2) = x_2^s$. Since the function g is convex, we have

$$\begin{aligned} \mathcal{V}(\mathbf{x}^s) &= p_1x_1^s + p_2x_2^s = p_1g(\theta_1) + p_2g(\theta_2) \geq g(\hat{\theta}) = \sqrt{2\theta_1^2 + 2\hat{\theta}(\hat{\theta} - \theta_1)} = \\ &= \sqrt{\theta_1^2 + \hat{\theta}^2 + (\hat{\theta} - \theta_1)^2} > \sqrt{\theta_1^2 + \hat{\theta}^2} \geq \sqrt{2\hat{\theta}\theta_1} = x_1^p = \mathcal{V}(\mathbf{x}^p), \end{aligned}$$

which says that the expected score under separation is greater than that under pooling. In this example, as more generally shown in Proposition 2, when the job market observes the grading rule applied to grade the student, the teacher sets up a grading rule that distinguishes between the types (provided, the score schedule $\mathbf{x}$ is convex).

Next, consider the case when the market cannot observe the grading rule applied, and suppose that it initially perceives that all the student types are being separated, i.e, $r_1 = \underline{r}$ and $r_2 = \overline{r}$. Under this circumstance, an individual teacher can "costlessly" reward the student with any ability signal, θ_1 or θ_2. If the teacher separates the types then the expected score level her student gets is as before:

$$\mathcal{V}(\mathbf{x}^s) = p_1\sqrt{2\theta_1^2} + p_2\sqrt{2\theta_1^2 + 2\theta_2(\theta_2 - \theta_1)}.$$

If the teacher pools the types for the highest ability reward of θ_2, then the expected score is equal to

$$\sqrt{2\theta_1\theta_2}.$$

Comparing the two expected-score expressions, we see that as $p_1 \to 1$ (i.e., as the low type becomes more likely) the pooling of types gives a higher expected score, while as $p_1 \to 0$ the separating of types does. By continuity, there is a threshold probability p^* such that for $p_1 > p^*$ pooling types yields a higher expected score than separating, and *vice versa* (at $p_1 = p^*$ the two rules yield the same expected score).

Coming back to the discussion of consistent grading standards held by the job market, for this two-type example we obtain that the separating grading rule is consistent when $p_1 < p^*$.

Partial internalization

In this two-type setting, it is also feasible to discuss the situation, when a teacher partially internalizes the effect her grading rule has on the perception of grading standards held by the market.

Suppose there are N teachers. If all the teachers set up the separating grading rule $\{\mathbf{x}^s, \mathbf{r}^s\}$ but one teacher, who pools the types for the highest grade, then the expected ability that the job market perceives from seeing the highest grade is given by

$$\tilde{\theta} = \frac{1}{N}\hat{\theta} + \frac{N-1}{N}\theta_2,$$

where $\hat{\theta} = p_1\theta_1 + p_2\theta_2$. The first term of the above expression is the effect an individual teacher's deviation from the separating grading rule has on the job market's perception of grading standards.

The separating grading rule is consistent if the expected effort under pooling, which is $\sqrt{2\theta_1\tilde{\theta}}$, is less than $\mathcal{V}(\mathbf{x}^s)$—the expected effort under separation. Suppose that for some N teachers it is the case, but also let $p_1 > p^*$ from the previous subsection hold. With more teachers arriving, i.e., with N increasing, the ability signal $\tilde{\theta}$ tends to θ_2 as an individual teacher's grading practice has a vanishing impact on the market perception of grades. Therefore, with $p_1 > p^*$, there must be a threshold number of teachers N^* such that for $N > N^*$ the separating grading rule can no longer remain consistent.

It should also be noted that the pooling grading rule can be consistent for any $N > 1$, i.e., if every teacher pools types, then an individual teacher may not find it optimal to separate types for its only partial effect on the market perception of grades. Hence, returning back to the analysis in the previous paragraph, the pooling grading rule can also persist even when N goes back below N^* as no individual teacher wants to separate the student types when everyone else pools them.

5 Discussion

Here, we discuss the main findings of the model(s) presented. The focus lies on the model with nominal grades of Section 3, which overlaps with the model with relative grades of Section 4 when the job market does not observe exact grading rules. Arguably, the scenario with unobservable grading rules is more widespread than that with observable grading rules, but at the same time the latter scenario serves as an important benchmark for discussion on policy applications.

Significantly, the grading patterns obtained here closely match those observed in practice lending credibility to our modeling framework of a teacher-student relationship. In this light, we also argue that our model(s) can be used as microfoundations for policy application purposes such as designing merit-pay programs for teachers.

Furthermore, if applied to job performance appraisal, the model with nominal grades can also shed light on the leniency bias with implications to the compression of ratings observed in the appraisal process.

5.1 Main Findings

Compression of grades and ratings

In general, the results obtained in this paper show that if there is a reason to think that the principal does not pay for or internalize the cost of rewarding the agent, then in attempt to elicit on average more effort the principal chooses to be more lenient with rewards than otherwise she would have been. In particular, the "no pooling at the top" property, generally observed in models with costly transfers, does not hold here.

Referring to the teacher-student relationship studied, if an individual teacher cannot credibly commit to her using the same grading standards the job market holds and the job market cannot distinguish among individual grading rules applied, then we face the situation when the teacher treats grades as costless rewards. As a result, the ubiquitous compression of grades or leniency in grading can be the expected–knowledge-maximizing outcome of the teacher's optimal grading: it aims to extract more effort from lower-ability students with the help of costless good grades. To put it differently, good grades are "the commons" that teachers exploit to their benefit. But as in any problem of the commons, we inevitably obtain the deterioration of the commons, which, in our case, takes the form of grade compression with implications to grade inflation, discussed below.

Compressed rewards are characteristic not only to grading.[13] The literature on job performance appraisal has long dealt with the phenomenon of compression of ratings. This phenomenon is about raters', e.g., managers', shallow differentiation of good from bad performance of their ratees, e.g., employees (Murphy & Cleveland (1995); also see Prendergast (1999) for economists' account on the issue). Since performance ratings are mostly used for salary administration purposes, a line manager can find himself in the position when he values his employees' effort but does not internalize the payroll cost resulting from his ratings given. Hence, this manager-employee relationship falls within our studied framework with costless rewards. As a result, the compression of ratings

[13] Johnson (1997) observes similarities of performance evaluation in academics, professional golf, airline industry, and others.

can be an optimal (employee-effort-maximizing) outcome for a manager, which, however, deserves a study of its own.

Mismatch between grades and abilities

Our results—Result 2, in particular—also offer an explanation why teachers of classes with less able students are more lenient in grading than others (see, e.g., Goldman & Widawski (1976)). As we argue, this can be an outcome of the optimal design of grading rules and not necessarily an outcome of some teachers' rent-seeking behavior, as sometimes is suggested (e.g., Johnson (1997)). Our model with nominal grades predicts that in classes with less able students the optimal grading rules are designed so that high grades are more easily attainable to elicit more effort from more numerous lower-ability students, resulting in a mismatch and low correlation between students' grades and their abilities. For instance, if the population of mathematics students contains more talented people than, say, the population of economics students, then we should observe stricter grading standards applied in mathematics classes and fewer good grades rewarded (as ample empirical evidence shows to be the case, which is explored in the following section).

Concerning the normative side of the differential grading standards discussed, there have been a number of papers proposing grade adjustment mechanisms (see, e.g., Johnson (1997)) in order to make grades more informative of students' actual abilities. Without going into the details of this literature, it is worth noting that there it is typically assumed that the true reason for differential grading standards lies with some personal features of the instructor (e.g., the adaptation level, unwillingness to spend office hours on dealing with students' complaints about low grades, etc.). Therefore, proposed grade adjustment mechanisms would attempt to correct for presumed instructor-specific factors failing to recognize the possible endogeneity of those factors, which could lead a mechanism astray from the projected goals.

Grade inflation

From our model(s) studied, we can distinguish two factors contributing toward grade inflation. First, teachers become more lenient with grading in response to shifts in distribution for student abilities toward the lower end of the ability space (Result 2 of this paper). Second, if an individual teacher finds that her grading practice does not affect the perception of the job market regarding the signaling value of grades, then the teacher tends to overuse good grades (see Results 3 and 4).

The two factors can both originate from the same source, namely, the expanding availability of education (see footnote 7 of this paper and, e.g., McKenzie & Tullock (1981, Ch. 17)). Due to an increasing number of educational institutions and study programs in recent decades, a larger number of study places has been offered resulting in more lower-ability applicants being enrolled (see footnote 5). Subsequently, this can lead to the emergence of the first factor discussed. Similarly, with more issuers of educational certificates, the grading rules applied by every issuer or teacher become less identifiable, resulting, correspondingly, in the emergence of the second factor.[14]

5.2 Policy Applications

The modeling framework presented can be used as microfoundations of a teacher-student relationship to analyze the implications of the introduction of merit-pay programs for teachers. In recent years, a number of such programs have been introduced in various countries to foster incentives for teachers in their endeavors to motivate more effort from their students (see, Lavy (2002, 2009); Atkinson *et al.* (2004); Lazear (2003)). Typically, these programs offer monetary bonuses to teachers if their students improve upon their previous performance (as measured by their scores achieved on standardized tests). The goals pursued by the developers of such programs—social planners—range from improving average performance (in most cases) to reducing the gap between poor and good performers (as, e.g., in the "No Child Left Behind" initiative in the US).

In terms of our modeling framework, the incentives for teachers set forth by merit-pay programs can have a direct effect on the form of the teacher's utility function, V, in (1). In the case of the "No Child Left Behind" program, where teachers are rewarded for a reduction in the gap between poor and good performers, the utility function V could turn concave since the teacher would start putting relatively more weight on the performance of lower-ability students. Then, with a concave utility function V, the model with nominal grades of Section 3[15] predicts that, compared with the case of linear utility, the gap between low- and high-ability students would diminish. However, this reduction would come from two directions, namely, from the teacher's demanding more effort from low-ability students and demanding less effort from high-ability students.[16] Hence, according

[14]The expanding availability of education is discussed as a cause of the grade inflation phenomenon in McKenzie & Tullock (1981, Ch. 17). Their hypothesized link is in the context of the demand for and supply of university openings: in response to a higher competition for students—due to the increasing number of university openings relative to the demand—universities engaged in lowering grading standards in order to attract more students. According to McKenzie & Tullock (1981), this practice eventually led to grade inflation.

[15]This model is perhaps more appropriate for modeling grading patterns in high schools, for ability-signaling concerns should be of a lesser magnitude among high-school students.

[16]With a concave utility function V, the optimal score allocations $\tilde{\mathbf{x}}(\theta)$ for θ in $[\underline{\theta}, \theta^*)$ are equal to

$$\tilde{\mathbf{x}}(\theta) = y_x^{-1}\left(\frac{f(\theta)\theta^2 V_x\left[\tilde{\mathbf{x}}(\theta)\right] y_x\left[\tilde{\mathbf{x}}(\theta^*)\right]}{f(\theta^*)\theta^{*2} V_x\left[\tilde{\mathbf{x}}(\theta^*)\right]}\right).$$

First, it has to be the case for the highest score levels that $\tilde{\mathbf{x}}(\theta^*) < \mathbf{x}(\theta^*)$, where $\mathbf{x}(\theta^*)$ is the pooling-interval score level in the linear-utility case. This is so because the ratio $V_x\left[\tilde{\mathbf{x}}(\theta)\right]/V_x\left[\tilde{\mathbf{x}}(\theta^*)\right]$ for $\theta < \theta^*$ is strictly greater than 1 and if $\tilde{\mathbf{x}}(\theta^*) \geq \mathbf{x}(\theta^*)$ then the whole score allocation schedule $\tilde{\mathbf{x}}$ is above the allocation schedule $\mathbf{x}$ of the linear case. But it is not possible because $\tilde{\mathbf{x}}$ would be the solution to the teacher's problem in the linear case, not $\mathbf{x}$. Next, the ratio $(V_x\left[\tilde{\mathbf{x}}(\theta)\right] y_x\left[\tilde{\mathbf{x}}(\theta^*)\right])/V_x\left[\tilde{\mathbf{x}}(\theta^*)\right]$ needs to be to the model, a negative externality from the introduction of this program can arise: high-ability students can be made get the same grades but for less effort. On the contrary, the gap between poor and good performers increases if the teacher's utility function turns convex—as a result, for instance, of the social planner implementing a merit-pay program that rewards for students' excellence only.

With the help our model(s), we can also offer an insight into the problem whether the university administration should restrict the teachers' choice of grading rules by imposing relative grading, i.e., grading on a curve. From the perspective of our model with relative grades, when the job market observes the grading rule applied by the teacher, an optimal grading rule is, actually, the one that perfectly screens student types. Then, imposing grading on a curve is superfluous since it does not bind. However, when the job market does not observe grading rules, the teacher faces a commitment problem of not overusing good grades, as we discussed before. In this event, grading on a curve would actually bind and could possibly fix the teacher's commitment problem. But then, the question is what goals the university administration pursues. If they coincide with the teacher's, i.e., maximizing student knowledge, then grading on a curve would be a desirable policy. But if the administration aims to maximize the expected wage of its students, assuming it is proportional to the ability signal inferred by the market, then the administration may want, as argued in Chan *et al.* (2007) and Ostrovsky & Schwarz (2010), to refrain from imposing grading on a curve and rather have grades compressed in order to disclose information about student abilities only coarsely.[17]

6 Empirical Evidence

Here we present empirical evidence in support of our Result 2—the lower the expectations the teacher holds about her students' abilities, the more lenient the grading rule she sets up.

In general, to test this theoretical prediction of the model, one would need university data such as student grades and their ability proxy (like their performance on university entry exams or Scholastic Aptitude Test [SAT] scores). Then, roughly speaking, one would compare grading patterns for classes with different student ability distributions and see if the prediction holds. However, there have been a number of empirical studies of the kind in the special literature of educational measurement (e.g., in academic periodicals such as the *Journal of Educational Measurement* or *Educational and Psychological Measurement*). Most importantly, those studies without exception report results that are

greater than $y_x\left[\mathbf{x}(\theta^*)\right]$ at least for some θ in $[\underline{\theta}, \theta^*)$, otherwise the score schedule in the linear case would do better than $\tilde{\mathbf{x}}$. Hence, if for some θ $\tilde{\mathbf{x}}(\theta) > \mathbf{x}(\theta)$ so it is for $\theta = \underline{\theta}$. Therefore, we obtain a reduction in the gap between the highest and lowest performances, but this reduction comes from both directions.

[17] For more discussion on and empirical implications of making academic transcripts more informative, see Bar *et al.* (2009).

fully in line with the model's predictions: fields with lower ability students studied as compared with those with higher ability students employ less stringent grading criteria. Even though many of those studies are comprehensive in empirical matters, they lack any rigorous theoretical explanation for this phenomenon. Their explanations mainly hinge on intuition or reference to similar phenomena from the adaptation-level theory in psychological literature. In what follows, we attempt to review in detail some of the empirical studies comparing grading standards over time and in different fields, and to show that our model proves helpful in explaining the empirical evidence observed.

Aiken (1963) is one of the first empirical studies that suggest that grading behavior is dictated by the quality of students in the current class and not by some absolute invariant standards. Aiken (1963) presents time-series evidence from the Woman's College of the University of North Carolina that could imply that, with more able students in a class (as measured by their SAT scores and high-school rankings), teachers tend to apply more stringent grading standards. As for the theoretical explanation for this finding, the study only briefly mentions that it conforms to the adaptation-level theory or central tendency phenomenon, which basically concerns the tendency of supervisors to evaluate the performance of people supervised in relative terms rather than in absolute ones.

A much more comprehensive study Goldman & Widawski (1976) first notes the weaknesses of previous studies on grading patterns because of their using the total grade point average (GPA) as the criterion of grading standards. As they rightly argue, GPAs are not perfectly comparable either over time or among individual students because of the possibly different composition of courses included to compute grade averages. To remedy that, Goldman & Widawski (1976) employ a between-subjects design aimed at making grade comparisons more effective. They compute an index of grading standards using pairwise comparisons of grades in 17 major fields at the University of California, Riverside, from a random sample of 475 students. In particular, they perform the comparison of grading standards in one class (say, psychology) against those in another class (say, biology) by computing the difference in average grades of only those students who took both classes. After obtaining differentials in grading standards between any two classes (from the 17 classes available in their study), they construct an index of grading standards for each class, which is an average of all the differentials between that particular class and the rest of the classes. Finally, they correlate the computed indices of grading standards with the average scores on the verbal and mathematical portions of the SAT test and high-school GPAs (*i.e.*, student ability proxies) of all the students majoring in those 17 classes. The main empirical finding in Goldman & Widawski (1976) is that the constructed index of grading standards correlates highly in a negative direction with student ability proxies. In other words, they conclude that professors in a field containing more able students tend to grade more stringently than do professors in fields with lower ability students. As a result, they find that the past performance and abilities of students account for only slightly more than 50 percent of the variance in grades, and suggest introducing some grade adjustment mechanism to make grades more informative of students' true abilities. Again, in giving an explanation for the empirical results obtained, they restrict their argument simply by making a reference to the adaptation-level theory that people are judged in comparison to their peers.

A similar study Goldman & Hewitt (1975), which along with presenting the empirical results (which draw the same conclusions about grading behavior as in the studies mentioned above), also provides a more elaborate theoretical explanation for the results obtained. The authors think that the antecedents (e.g., student ability levels, work habits, etc.) and consequences (grading standards) of college grading are inextricably tied together by a personal characteristic of college instructors. This characteristic is the phenomenon of adaptation level, and it is so pervasive among college instructors and perhaps people in general, Goldman & Hewitt (1975) continue, as to be considered an almost inevitable factor in the college grading process. Consequently, through that personal characteristic link, grading standards would be partly determined by the ability level of the student population. However, along the lines of our model developed above, this personal characteristic, as envisaged by Goldman & Hewitt (1975), is not some intrinsic feature of human behavior but rather the outcome of optimal behavior.

A decade later, Strenta & Elliott (1987) replicated the study of Goldman & Widawski (1976) using data from a different institution, Dartmouth College, just to find that the differential grading standards exist in the same magnitude and in roughly the same order. Therefore, Strenta & Elliott (1987) argue that it remains the case that students with higher SAT scores tend to major in fields with more rigorous grading standards, and that factors attracting more talented students result in their being graded harder. (However, we would argue for the reverse direction of causation: since some fields attract more talented students, professors in those fields will grade their students more stringently, which is optimal in order to extract more effort.) As in previous studies, Strenta & Elliott (1987) argue that these differential grading standards serve to attenuate the correlation between the GPAs and SAT scores of the students, and they also show that the correlation increases sizably if GPAs are adjusted by accounting for differences in departmental grading standards. Finally, a similar study conducted at Duke University (Johnson (2003)) confirmed the conclusions about systematic differences in grading standards from the previous studies.

7 Conclusion

In this paper, we consider a teacher-student relationship as a special type of an agency problem featuring costless rewards. Our theoretical predictions offer a good match to grading patterns empirically observed both from the static and dynamic perspectives.

This allows us to suggest that the chosen modeling framework is appropriate to analyze a teacher-student relationship.

Appendix A. Proof of Proposition 1

Here, we solve the teacher's utility maximization problem (3)–(6), namely, we look for the grading rule $\{\mathbf{x}, \mathbf{r}\} \in \mathcal{C}^1_{\mathbf{x}} \times \mathcal{C}^1_{\mathbf{r}}$ that maximizes (3) subject to (4)–(6).

As it is standard for this type of problems, we start with reducing the maximization problem by singling out the set of constraints that bind in the optimum. First, assuming, for convenience, that the teacher never finds it optimal to set a zero allocation for any student type, in the optimum the individual rationality constraint of the lowest type is binding

$$U(\underline{\theta}, \underline{\theta}) = 0.$$

Second, the set of adjacent incentive-compatibility constraints needs to be downward binding, from which the student's utility levels for types $\theta \in (\theta, \overline{\theta}]$ in the teacher's optimum are equal to

$$U(\theta, \theta) = -\int_{\underline{\theta}}^{\theta} C_\theta(\mathbf{x}(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}. \tag{23}$$

When the above constraints hold and we have a monotonously increasing score schedule $\mathbf{x}$ (which is a necessary condition for incentive compatibility), then, by the single crossing property the rest of constraints will also hold.

Next, we observe that since grades are costless for the teacher to reward, in the optimum it must be that the highest ability type gets the highest grade, i.e., $\mathbf{r}(\overline{\theta}) = \overline{r}$. Then, we make the following conjecture.

Conjecture A.1 *In the solution to (3)–(6),* $\mathbf{r}(\theta) = \overline{r}$ *only for* $\theta = \overline{\theta}$*.*

To put it differently, we conjecture that there is no pooling of types at the top. Now, from (23) for $\theta = \overline{\theta}$ we get

$$\overline{r} - C(\mathbf{x}(\overline{\theta}),\overline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} C_\theta(\mathbf{x}(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta} = 0, \tag{24}$$

which combines all the constraints into one and eliminates the grade schedule from the maximization problem (if (24) is satisfied, so are all the other constraints by choosing the "right" grade schedule).

Next, we set up the Lagrangean of the reduced maximization problem:

$$\mathcal{L}(\mathbf{x}, \lambda) = \int_{\underline{\theta}}^{\overline{\theta}} \mathbf{x}(\theta) f(\theta) d(\theta) + \lambda[\overline{r} - C(\mathbf{x}(\overline{\theta}),\overline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} C_\theta(\mathbf{x}(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}].$$

The first-order conditions (FOCs) are

$$f(\overline{\theta}) + \lambda[-C_x(\mathbf{x}(\overline{\theta}),\overline{\theta}) + C_{x\theta}(\mathbf{x}(\overline{\theta}),\overline{\theta})] = 0 \tag{25}$$

with respect to allocation $\mathbf{x}(\overline{\theta})$, and

$$f(\theta) + \lambda C_{x\theta}(\mathbf{x}(\theta),\theta) = 0. \tag{26}$$

with respect to allocations $\mathbf{x}(\theta)$, $\theta \in [\underline{\theta}, \overline{\theta})$. These FOCs together with reduced constraint (24) characterize the optimal score schedule $\mathbf{x}$ provided Conjecture A.1 holds.

Combining through the Lagrange multiplier λ the first-order condition (25) with (26), we get

$$\frac{f(\overline{\theta})}{f(\theta)} = \frac{C_x(\mathbf{x}(\overline{\theta}),\overline{\theta}) - C_{x\theta}(\mathbf{x}(\overline{\theta}),\overline{\theta})}{-C_{x\theta}(\mathbf{x}(\theta),\theta)},$$

which should hold for any type θ. But if we take the limit $\theta \to \overline{\theta}$ on both sides, we find that the left-hand side converges to 1, while the right-hand side — to something strictly greater than one:

$$\frac{C_x(\mathbf{x}(\overline{\theta}),\overline{\theta})}{-C_{x\theta}(\lim_{\theta\to\overline{\theta}} \mathbf{x}(\theta),\overline{\theta})} + \frac{C_{x\theta}(\mathbf{x}(\overline{\theta}),\overline{\theta})}{C_{x\theta}(\lim_{\theta\to\overline{\theta}} \mathbf{x}(\theta),\overline{\theta})} \geq \frac{C_x(\mathbf{x}(\overline{\theta}),\overline{\theta})}{-C_{x\theta}(\lim_{\theta\to\overline{\theta}} \mathbf{x}(\theta),\overline{\theta})} + 1 > 1.$$

Hence, there cannot exist a monotonous score schedule $\mathbf{x}$ that satisfies the first-order conditions and constraint (24). In other words, there is no shadow price λ that can balance all the incentives and screen types at the top. Therefore, Conjecture A1 does not hold implying that there must be some pooling of types for the highest reward.

Therefore, we proceed by pooling types θ which are subject to the uniform allocation with the highest grade of $\overline{r}$. Let θ^* denote the starting value of the pooling interval $[\theta^*, \overline{\theta}]$, and the constraint equivalent to (24) becomes

$$\overline{r} - C(\mathbf{x}(\theta^*),\theta^*) + \int_{\underline{\theta}}^{\theta^*} C_\theta(\mathbf{x}(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta} = 0. \tag{27}$$

The Lagrangean now takes the form of

$$\mathcal{L}(\mathbf{x}, \lambda) = \int_{\underline{\theta}}^{\theta^*} \mathbf{x}(\theta) f(\theta) d(\theta) + (1 - F(\theta^*))\mathbf{x}(\theta^*) + \lambda[\overline{r} - C(\mathbf{x}(\theta^*),\theta^*) + \int_{\underline{\theta}}^{\theta^*} C_\theta(\mathbf{x}(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}].$$

The FOCs are

$$f(\theta^*) + (1 - F(\theta^*)) + \lambda(-C_x(\mathbf{x}(\theta^*),\theta^*) + C_{x\theta}(\mathbf{x}(\theta^*),\theta^*)) = 0 \tag{28}$$

with respect to allocation $\mathbf{x}(\theta^*)$, and, as before,

$$f(\theta) + \lambda C_{x\theta}(\mathbf{x}(\theta),\theta) = 0 \tag{29}$$

with respect to allocations $\mathbf{x}(\theta)$, $\theta \in [\underline{\theta}, \theta^*)$. Combining the two conditions, we get

$$\frac{f(\theta^*) + (1 - F(\theta^*))}{f(\theta)} = \frac{C_x(\mathbf{x}(\theta^*),\theta^*) - C_{x\theta}(\mathbf{x}(\theta^*),\theta^*)}{-C_{x\theta}(\mathbf{x}(\theta),\theta)}.$$

Before deriving the condition for the starting point of the pooling interval θ^*, we make the following observation. In the solution, the score schedule $\mathbf{x}$ needs to be continuous at the starting point of the pooling interval. If it were not, then the reward schedule $\mathbf{r}$ would also be discontinuous (otherwise, the grading rule would not be incentive compatible). But since grades are costless for the teacher, then at no cost she can improve her utility by tilting up the segment of score allocations to the left from the discontinuity point and accordingly adjusting the grade allocations to meet the incentive compatibility constraints. Hence, the score schedule $\mathbf{x}$ cannot be discontinuous at the starting point of the pooling interval.[18]

Taking the limit $\theta \to \theta^*$ on both sides and using the continuity of $\mathbf{x}$ at θ^* we get the condition for the starting point of the pooling interval $[\theta^*, \overline{\theta}]$:

$$\frac{1 - F(\theta^*)}{f(\theta^*)} = \frac{C_x(\mathbf{x}(\theta^*),\theta^*)}{-C_{x\theta}(\mathbf{x}(\theta^*),\theta^*)},$$

or, given the functional assumption that the effort cost function C is separable in score and type, $C(x, \theta) = y(x)/\theta$, this condition becomes

$$\frac{1 - F(\theta^*)}{f(\theta^*)} = \theta^*. \tag{30}$$

Since there may be no type θ in $\left[\underline{\theta}, \overline{\theta}\right]$ for which condition (30) holds[19], then the starting value of the "pooling at the top" interval is defined as

$$\theta^* = \min\{\theta : (1 - F(\theta))/(\theta f(\theta)) \leq 1, \theta \in \Theta\}, \tag{31}$$

which is (7) in Proposition 1. Note that the expression $(1-F(\theta))/(f(\theta)\theta)$ is monotonically decreasing in θ due to the monotone hazard rate assumption so that the pooling-interval starting point θ^* is uniquely determined.

Suppose that $\theta^* > \underline{\theta}$. Denote the score-grade allocation for every type θ in $\left[\theta^*, \overline{\theta}\right]$ by

[18] The argument given also depends on the fact that the score allocations to the left from θ^* are strictly monotonous, which immediately follows from the monotone hazard rate condition and the observation that $\theta f(\theta) < 1 - F(\theta)$ for $\theta \in [\underline{\theta}, \theta^*)$ following from the pooling-interval condition.

[19] The pooling interval comprises the whole type space if, for instance, student types θ are uniformly distributed with $\underline{\theta} \geq \overline{\theta}/2$.

$(\mathbf{x}(\theta^*), \overline{r})$, where the score allocation $\mathbf{x}(\theta^*)$ needs still to be determined. From (28) and (30) we express the Lagrange multiplier λ to be equal to

$$\lambda = \frac{f(\theta^*)\theta^{*2}}{y_x(\mathbf{x}^*(\theta^*))}. \tag{32}$$

Plugging the above expression for λ into remaining first-order conditions (29), we get that for every θ in $[\underline{\theta}, \theta^*)$ the optimal score allocation $\mathbf{x}(\theta)$ is equal to

$$\mathbf{x}(\theta) = y_x^{-1}\left(\frac{f(\theta)\theta^2 y_x(\mathbf{x}(\theta^*))}{f(\theta^*)\theta^{*2}}\right), \tag{33}$$

which is (8) in Proposition 1.

Finally, the highest score allocation, $\mathbf{x}(\theta^*)$, can be determined from the constraint

$$\overline{r} - C(\mathbf{x}(\theta^*), \theta^*) + \int_{\underline{\theta}}^{\theta^*} C_\theta(\mathbf{x}(\theta), \theta)d\theta = 0, \tag{34}$$

after plugging in the expression for $\mathbf{x}(\theta)$ from (33), giving (10) in Proposition 2.

The constraint that the score schedule $\mathbf{x}$ be non-decreasing is met, which follows from (30) and the monotone hazard rate assumption.[20]

The optimal grade allocations $\mathbf{r}(\theta)$ for θ in $[\underline{\theta}, \theta^*)$ are found from (23) and are equal to

$$\mathbf{r}(\theta) = C(\mathbf{x}(\theta), \theta) - \int_{\underline{\theta}}^{\theta} C_\theta(\mathbf{x}(\tilde{\theta}), \tilde{\theta})d\tilde{\theta}, \tag{35}$$

which is (9) in Proposition 1, concluding the solution to the optimization problem (3)–(6).

Appendix B. Proof of Result 2

With reference to the pooling-interval condition (7), define $g_i(\theta) = (1 - F_i(\theta))/(\theta f_i(\theta))$, then, the starting points of the pooling intervals are $\theta_i^* = \min\{\theta : g_i(\theta) \leq 1, \theta \in \Theta\}$, $i = 1, 2$. Since the likelihood ratio order implies the hazard rate order (see Shaked & Shanthikumar (1994)), which is $f_1(\theta)/(1 - F_1(\theta)) \leq f_2(\theta)(1 - F_2(\theta))$ for every θ, it immediately follows that $g_1(\theta) \geq g_2(\theta)$, leading to $\theta_1^* \geq \theta_2^*$. Hence, we have $\mathbf{r}_2(\theta) = \overline{r} \geq \mathbf{r}_1(\theta)$ for $\theta \in \left[\theta_2^*, \overline{\theta}\right]$.

Next, consider the optimal score allocations $\mathbf{x}_i(\theta)$, $i = 1, 2$, for types θ in $[\underline{\theta}, \theta_2^*)$. Denote the Lagrange multipliers from the two optimization problems by λ_1 and λ_2, defined by (32) of Appendix A, respectively. Divide the first-order conditions for $\mathbf{x}_1$ and $\mathbf{x}_2$ in

[20] From equation (30) it follows that $f(\theta)\theta/(1 - F(\theta)) < 1$ for θ in $[\underline{\theta}, \theta^*)$, and from the monotone hazard rate: $f'(\theta) > -f^2(\theta)/(1 - F(\theta))$. The two properties ensure that the derivative of (33), $\partial\mathbf{x}/\partial\theta$, is positive.

(29) of Appendix A to obtain for any θ in $[\underline{\theta}, \theta_2^*)$

$$\frac{y_x(\mathbf{x}_2(\theta))}{y_x(\mathbf{x}_1(\theta))} = \frac{\lambda_1}{\lambda_2}\frac{f_2(\theta)}{f_1(\theta)},$$

which also holds at $\theta = \theta_2^*$ by the continuity of the score schedule $\mathbf{x}_2$ at $\theta = \theta_2^*$ as argued in Appendix A.

Since the highest score $\mathbf{x}_2(\theta_2^*)$ in the second (less able) class must be at least as large as $\mathbf{x}_1(\theta_2^*)$, which stems from the second teacher's incentive to expand the pooling interval even further (otherwise the score schedule $\mathbf{x}_2$ could be improved upon by tilting it up), then at $\theta = \theta_2^*$ the left-hand side of the above expression is greater than or equal to 1, and so is the right-hand side. Due to the decreasing likelihood ratio $f_2(\theta)/f_1(\theta)$, the right-hand side stays greater than 1 for any θ in $[\underline{\theta}, \theta_2^*)$, and so does the left-hand side, implying that $\mathbf{x}_2(\theta) \geq \mathbf{x}_1(\theta)$ for every θ in $[\underline{\theta}, \theta_2^*)$, which subsequently leads to $\mathbf{r}_2(\theta) \geq \mathbf{r}_1(\theta)$ from (9) in Proposition 1.

Appendix C. Proof of Proposition 2

Define the teacher's expected utility $\mathcal{V} : \mathcal{C}_{\mathbf{x}}^1 \to \mathbb{R}$ from implementing a score schedule $\mathbf{x}$ by

$$\mathcal{V}(\mathbf{x}) = \int_{\underline{\theta}}^{\overline{\theta}} \mathbf{x}(\theta) f(\theta) d\theta.$$

Let the score schedule $\mathbf{x}^*$ with the grade schedule $\mathbf{r}$ imposed by (16) solve the teacher's problem (13)–(15) and suppose that there is a non-empty pooling interval $[\theta', \theta'']$ at which $\mathbf{x}^*(\theta) = x'$ and $\mathbf{r}(\theta) = r'$ (the arguments below are also valid if we consider a half-open or open pooling interval). We need to consider two cases 1) $\theta' = \underline{\theta}$ and 2) $\theta' > \underline{\theta}$. In each case it is sufficient to restrict attention to those score allocations that satisfy the optimality conditions: the binding individual rationality constraint of the lowest-ability type and the set of downward-binding incentive compatibility constraints, respectively.

In the first case, 1) $\theta' = \underline{\theta}$, the score allocation $\mathbf{x}^*(\theta) = x'$ with grade r' for all θ in $[\underline{\theta}, \theta'']$ need to satisfy the binding individual rationality constraint of the lowest-ability type:

$$\theta^{\mathbf{r}}(r') - \frac{y(x')}{\underline{\theta}} = 0,$$

or

$$x' = y^{-1}[\theta^{\mathbf{r}}(r')\underline{\theta}],$$

where the function y^{-1} is the inverse of y and $\theta^{\mathbf{r}}$ is the ability type inferred as given by

(11). The teacher's expected utility from implementing the score schedule $\mathbf{x}^*$ is given by

$$\mathcal{V}(\mathbf{x}^*) = F(\theta'')y^{-1}[\theta^{\mathbf{r}}(r')\underline{\theta}] + \int_{\theta''}^{\overline{\theta}} \mathbf{x}^*(\theta)f(\theta)d\theta.$$

As an alternative to the score schedule $\mathbf{x}^*$, consider the following score schedule $\hat{\mathbf{x}}$: for θ in $[\underline{\theta}, \theta'']$ set (distinct and incentive compatible) performance allocations $\hat{\mathbf{x}}(\theta) = y^{-1}[(\theta^2 + \underline{\theta}^2)/2]$, as in (17), and for θ in $(\theta'', \overline{\theta}]$ — $\hat{\mathbf{x}}(\theta) = \mathbf{x}^*(\theta)$. The case of interest is the situation when the monotonicity of the new score schedule $\hat{\mathbf{x}}$ is preserved. Otherwise, when the monotonicity not preserved, i.e., if $\hat{\mathbf{x}}(\theta'') > \hat{\mathbf{x}}(\theta) = \mathbf{x}^*(\theta)$ for some $\theta > \theta''$, the teacher can increase her expected utility by simply setting $\hat{\mathbf{x}}(\theta) = \hat{\mathbf{x}}(\theta'')$ for all $\theta > \theta''$ such that $\hat{\mathbf{x}}(\theta) < \hat{\mathbf{x}}(\theta'')$ and leave the remaining allocations intact. The teacher's expected utility from implementing the score schedule $\hat{\mathbf{x}}$ is equal to

$$\mathcal{V}(\hat{\mathbf{x}}) = \int_{\underline{\theta}}^{\theta''} \hat{\mathbf{x}}(\theta)f(\theta)d\theta + \int_{\theta''}^{\overline{\theta}} \mathbf{x}^*(\theta)f(\theta)d\theta.$$

The second terms of $\mathcal{V}(\mathbf{x}^*)$ and $\mathcal{V}(\hat{\mathbf{x}})$ are identical, and so any difference in the utilities needs to come from the difference in the first terms. Since the new performance allocation schedule $\hat{\mathbf{x}}$ is convex on the restriction $[\underline{\theta}, \theta'']$—the condition of Proposition 2—then by Jensen's inequality

$$\begin{aligned} \int_{\underline{\theta}}^{\theta''} \hat{\mathbf{x}}(\theta)f(\theta)d(\theta) &\geq F(\theta'')\hat{\mathbf{x}}(\int_{\underline{\theta}}^{\theta''} \theta f(\theta)d(\theta)/F(\theta'')) = F(\theta'')\hat{\mathbf{x}}(\theta^{\mathbf{r}}(r')) \\ &= F(\theta'')y^{-1}[(\theta^{\mathbf{r}}(r')^2 + \underline{\theta}^2)/2] > F(\theta'')y^{-1}(\theta^{\mathbf{r}}(r')\underline{\theta}), \end{aligned}$$

which is equal to the first term of $\mathcal{V}(\mathbf{x}^*)$, and where the last inequality stems from the fact that the arithmetic average is greater than the geometric one and y^{-1} is strictly increasing.

Hence, instead of pooling ability types at the bottom the teacher can do better by screening them since $\mathcal{V}(\hat{\mathbf{x}}) > \mathcal{V}(\mathbf{x}^*)$.

In the second case, $\theta' > \underline{\theta}$, the allocation $\mathbf{x}^*(\theta) = x'$ for all θ in $[\theta', \theta'']$ together with grade r' need to satisfy the downward binding incentive compatibility constraint of the ability type θ', which can be expressed as

$$\theta^{\mathbf{r}}(r') - \frac{y(x')}{\theta'} = \theta^{\mathbf{r}}(r'') - \frac{y(x'')}{\theta'},$$

where the allocation (x'', r'') is the best alternative to type θ'. Since the teacher screens the types in $[\underline{\theta}, \theta')$ and the optimal way of doing it is as in (17)—otherwise, the score schedule $\mathbf{x}^*$ is not optimal—then the score allocation $x'' = \sup_{\theta<\theta'} \mathbf{x}^*(\theta) = y^{-1}[(\theta'^2 + \underline{\theta}^2)/2]$ and the ability signal $\theta^{\mathbf{r}}(r'') = \sup_{\theta<\theta'} \theta = \theta'$. The optimal score allocation x' can, accordingly, be expressed as

$$x' = y^{-1}\left[\theta'\theta^{\mathbf{r}}(r') - \frac{\theta'^2 - \underline{\theta}^2}{2}\right],$$

and the resulting expected utility to the teacher from implementing the grading rule $\mathbf{x}^*$ is equal to

$$\mathcal{V}(\mathbf{x}^*) = \int_{\underline{\theta}}^{\theta'} \mathbf{x}^*(\theta) f(\theta) d\theta + (F(\theta'') - F(\theta'))x' + \int_{\theta''}^{\overline{\theta}} \mathbf{x}^*(\theta) f(\theta) d\theta.$$

Similarly to the previous case, consider the following grading rule $\hat{\mathbf{x}}$: for θ in $[\theta', \theta'']$ set performance allocations $\hat{\mathbf{x}}(\theta) = y^{-1}[(\theta^2 + \underline{\theta}^2)/2]$ and for θ in $[\underline{\theta}, \theta') \cup (\theta'', \overline{\theta}]$ — $\hat{\mathbf{x}}(\theta) = \mathbf{x}^*(\theta)$. (The monotonicity of the grading rule $\hat{\mathbf{x}}$ is preserved on the restriction $[\underline{\theta}, \theta'']$ by the construction of $\hat{\mathbf{x}}(\theta)$ for θ in $[\theta', \theta'']$, and regarding the monotonicity over $(\theta'', \overline{\theta}]$ the same argument as in the first case studied above applies.) The teacher's expected utility from implementing $\hat{\mathbf{x}}$ is equal to

$$\mathcal{V}(\hat{\mathbf{x}}) = \int_{\underline{\theta}}^{\theta'} \mathbf{x}^*(\theta) f(\theta) d\theta + \int_{\theta'}^{\theta''} \hat{\mathbf{x}}(\theta) dF(\theta) + \int_{\theta''}^{\overline{\theta}} \mathbf{x}^*(\theta) f(\theta) d\theta.$$

Given the condition that $\hat{\mathbf{x}}(\theta)$ is convex for θ in $[\theta', \theta'']$, by Jensen's inequality we have for the second term of $\mathcal{V}(\hat{\mathbf{x}})$ that

$$\begin{aligned}
& \int_{\theta'}^{\theta''} \hat{\mathbf{x}}(\theta) dF(\theta) \geq (F(\theta'') - F(\theta'))\hat{\mathbf{x}}(\theta^{\mathbf{r}}(r')) = \\
= \ & (F(\theta'') - F(\theta'))y^{-1}[(\theta^{\mathbf{r}}(r')^2 + \underline{\theta}^2)/2] > \\
> \ & (F(\theta'') - F(\theta'))y^{-1}[\theta'\theta^{\mathbf{r}}(r') - (\theta'^2 - \underline{\theta}^2)/2],
\end{aligned}$$

which is the second term of $\mathcal{V}(\mathbf{x}^*)$, and where the last inequality stems from the fact that $\theta^{\mathbf{r}}(r') > \theta'$ and y^{-1} is strictly increasing. From this derivation, we have again that $\mathcal{V}(\hat{\mathbf{x}}) > \mathcal{V}(\mathbf{x}^*)$, which concludes the proof of Proposition 2 that pooling student types cannot be optimal for the teacher when the score allocation $\mathbf{x}$ in (17) is convex.

References

Aiken, Jr., Lewis R. 1963. The Grading Behavior of a College Faculty. *Educational and Psychological Measurement*, **23**(2), 319–322.

Arcidiacono, Peter. 2004. Ability sorting and the returns to college major. *Journal of Econometrics*, **121**(1-2), 343–375.

Atkinson, Adele, Burgess, Simon, Croxson, Bronwyn, Gregg, Paul, Propper, Carol, Slater, Helen, & Wilson, Deborah. 2004. *Evaluating the Impact of Performance-related Pay for Teachers in England*. CMPO working paper No. 04/113.

Babcock, Philip S., & Marks, Mindy. 2010 (April). *The Falling Time Cost of College: Evidence from Half a Century of Time Use Data*. NBER Working Paper No. 15954.

Bar, Talia, Kadiyali, Vrinda, & Zussman, Aaf. 2009. Grade Information and Grade Inflation: The Cornell Experiment. *Journal of Economic Perspectives*, **23**(3), 93–108.

Bedard, Kelly. 2001. Human Capital versus Signaling Models: University Access and High School Dropouts. *Journal of Political Economy*, **109**(4), 749–775.

Besley, Timothy, & Ghatak, Maitreesh. 2008. Status Incentives. *American Economic Review*, **98**(2), 206–211.

Chan, William, Hao, Li, & Suen, Wing. 2007. A Signaling Theory of Grade Inflation. *International Economic Review*, **48**, 1065–1090.

Dickson, Vaughan A. 1984. An Economic Model of Faculty Grading Practices. *The Journal of Economic Education*, **15**(3), 197–203.

Dubey, Pradeep, & Geanakoplos, John. 2010. Grading Exams: 100, 99, 98, ... or A, B, C? *Games and Economic Behavior*, **69**, 72–94.

Goldman, Roy D., & Hewitt, Barbara Newlin. 1975. Adaptation-Level as an Explanation for Differential Standards in College Grading. *Journal of Educational Measurement*, **12**(3), 149–161.

Goldman, Roy D., & Widawski, Mel H. 1976. A Within-Subjects Technique for Comparing College Grading Standards: Implications in the Validity of the Evaluation of College Achievement. *Educational and Psychological Measurement*, **36**(2), 355–359.

Guesnerie, Roger, & Laffont, Jean-Jacques. 1984. A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm. *Journal of Public Economics*, **25**(3), 329–369.

Johnson, Valen E. 1997. An Alternative to Traditional GPA for Evaluating Student Performance. *Statistical Science*, **12**(4), 251–269.

Johnson, Valen E. 2003. *Grade Inflation: A Crisis in College Education*. Springer.

Krishna, Vijay. 2002. *Auction Theory*. Academic Press.

Kuh, G., & Hu, S. 1999. Unraveling the Complexity of the Increase in College Grades from the Mid-1980s to the Mid-1990s. *Educational Evaluation and Policy Analysis*, **20**, 297–320.

Lavy, Victor. 2002. Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement. *Journal of Political Economy*, **110**(6), 1286–1317.

Lavy, Victor. 2009. Performance Pay and Teachers' Effort, Productivity, and Grading Ethics. *American Economic Review*, **99**(5), 1979–2011.

Lazear, Edward. 2003. Teacher Incentives. *Swedish Economic Policy Review*, **10**, 179–2003.

Levin, Jonathan. 2003. Relational Incentive Contracts. *American Economic Review*, **93**(3), 835–857.

MacLeod, W. Bentley. 2003. Optimal Contracting with Subjective Evaluation. *American Economic Review*, **93**(1), 216–240.

McKenzie, Richard B., & Tullock, Gordon. 1981. *The New World of Economics*. Homewood, Illinois: Richard D. Irwin.

Mirrlees, James A. 1971. An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies*, **38**(2), 175–208.

Moldovanu, Benny, Sela, Aner, & Shi, Xianwen. 2007. Contests for Status. *Journal of Political Economy*, **115**(2), 338–363.

Murphy, Kevin R., & Cleveland, Jeanette. 1995. *Understanding Performance Appraisal: Social, Organizational, and Goal-based Perspectives*. SAGE.

Ostrovsky, Michael, & Schwarz, Michael A. 2010. Information Disclosure and Unraveling in Matching Markets. *American Economic Journal: Microeconomics*, **2**(2), 34–63.

Prendergast, Canice. 1999. The Provision of Incentives in Firms. *Journal of Economic Literature*, **37**(1), 7–63.

Sabot, Richard, & Wakeman-Linn, John. 1991. Grade Inflation and Course Choice. *Journal of Economic Perspectives*, **5**, 159–170.

Shaked, Moshe, & Shanthikumar, J. George. 1994. *Stochastic Orders and Their Applications*. Acedemic Press, Inc.

Spence, Michael. 1973. Job Market Signaling. *Quarterly Journal of Economics*, **87**(3), 355–374.

Strenta, A. Christopher, & Elliott, Rogers. 1987. Differential Grading Standards Revisited. *Journal of Educational Measurement*, **24**(4), 281–291.

Wilson, Bradford P. 1999. The Phenomenon of Grade Inflation in Higher Education. *National Forum*, **Fall**, 38–41.